NPS-AM-10-049

EXCERPT FROM THE

PROCEEDINGS

OF THE

SEVENTH ANNUAL ACQUISITION RESEARCH SYMPOSIUM WEDNESDAY SESSIONS VOLUME I

Acquisition Research

Creating Synergy for Informed Change

May 12 - 13, 2010

Published: 30 April 2010

Prepared for: Naval Postgraduate School, Monterey, California 93943

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
MAY 2010		**00-00-2010 to 00-00-2010**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Towards Real-time Program Awareness via Lexical Link Analysis	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Postgraduate School,Information Sciences Department,1411 Cunningham Road, Root Hall 229,Monterey,CA,93943**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
7th Annual Acquisition Research Symposium to be held May 12-13, 2010 in Monterey, California

14. ABSTRACT

DoD acquisition is an extremely complex system, comprised of myriad stakeholders processes, people, activities, and organizations in an effort to provide the most useful capabilities to warfighters at the best possible value to the government. This effort is being accomplished by acquisition analysts who despite years of experience are encumbered by mountains of available data. To assist the analyst, we consider that the cognitive interface between decision-makers and a complex system may be expressed in a range of terms or ?features,? i.e., specific vocabulary to describe attributes. This offers the opportunity to more easily compare two competing technologies, which, in turn, may be compared to the Navy warfighter requirements. This effort can allow decision-makers to become aware of what programs, systems, and specific features are available for acquisition and how well they match warfighter?s needs and requirements with greater effect and immediacy?possibly in real-time. We present a data-driven automation method, namely, Lexical Link Analysis (LLA), to facilitate and automate acquisition system self-awareness.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**58**	

The research presented at the symposium was supported by the Acquisition Chair of the Graduate School of Business & Public Policy at the Naval Postgraduate School.

To request Defense Acquisition Research or to become a research sponsor, please contact:

NPS Acquisition Research Program
Attn: James B. Greene, RADM, USN, (Ret.)
Acquisition Chair
Graduate School of Business and Public Policy
Naval Postgraduate School
555 Dyer Road, Room 332
Monterey, CA 93943-5103
Tel: (831) 656-2092
Fax: (831) 656-2253
E-mail: jbgreene@nps.edu

Copies of the Acquisition Sponsored Research Reports may be printed from our website www.acquisitionresearch.net

Towards Real-time Program Awareness via Lexical Link Analysis

Ying Zhao—Dr. Ying Zhao is a Research Associate Professor at the Naval Postgraduate School (NPS). Dr. Zhao joined NPS in May 2009. Her research is focused on knowledge management approaches, e.g., data text/mining such as lexical link analysis, search and visualization for system self-awareness, decision-making and collaboration. She received her PhD in mathematics from MIT and co-founded Quantum Intelligence, Inc. She has been Principal Investigator (PI) for six DoD Small Business Innovation Research (SBIR) awarded contracts. She is a co-author of two patents in knowledge pattern search from networked agents, fusion and visualization for multiple anomaly detection systems.

Dr. Ying Zhao
Research Associate Professor
Naval Postgraduate School Information Sciences Department
1411 Cunningham Road, Root Hall 229
Monterey, CA 93943-5000
E-mail: yzhao@nps.edu; Phone: 831-656-3789

Shelley Gallup—Dr. Shelley Gallup is a Research Associate Professor at the Naval Postgraduate School's Department of Information Sciences, and Director of Distributed Information and Systems Experimentation (DISE). Dr. Gallup has a multi-disciplinary science, engineering and analysis background, including microbiology, biochemistry, space systems, international relations, strategy and policy and systems analysis. He returned to academia after retiring from naval service in 1994 and received his PhD in engineering management from Old Dominion University in 1998. Dr. Gallup joined NPS in 1999, bringing his background in systems analysis, naval operations, military systems and experimental methods first to the Fleet Battle Experiment series (1999-2002) and then to the FORCEnet experimentation in the Trident Warrior series of experiments (2003-present).

Dr. Shelley P. Gallup
Research Associate Professor
Naval Postgraduate School; Director
Distributed Information and Systems Experimentation (DISE) Group
Information Sciences Department
1411 Cunningham Road, Root Hall 103D
Monterey, CA 93943-5000
E-mail: spgallup@nps.edu; Phone: 831-656-1040

Doug MacKinnon—Dr. Doug MacKinnon is a Research Associate Professor at the Naval Postgraduate School (NPS). Dr. MacKinnon led an NPS research team to assess new MDA, spiral-1 technologies being fielded by PEO C4I developing original decision matrix structures and metrics structures to leverage the new technology. He has also led the assessment of TPED (Tasking, Planning, Exploitation, and Dissemination) process during field experiments Empire Challenge 2008 and 2009 (EC08/09). He holds a PhD from Stanford University, conducting theoretic and field research in Knowledge Management (KM). He has served as the program manager for two major government projects of over $50 million each, implementing new technologies while reducing manpower requirements. He has served over 20 years as a Naval Surface Warfare Officer, amassing over 8 years at sea, serving in four US Navy warships with five major, underway deployments.

Dr. Douglas J. MacKinnon
Research Associate Professor
Naval Postgraduate School
Information Sciences Department
Graduate School of Operational and Information Sciences
1411 Cunningham Road, Glasgow West, Room 3008
Monterey, CA 93943-5000
E-mail: djmackin@nps.edu; Phone: 831-656-1005

Abstract

DoD acquisition is an extremely complex system, comprised of myriad stakeholders, processes, people, activities, and organizations in an effort to provide the most useful capabilities to warfighters at the best possible value to the government. This effort is being accomplished by acquisition analysts who despite years of experience are encumbered by mountains of available data. To assist the analyst, we consider that the cognitive interface between decision-makers and a complex system may be expressed in a range of terms or "features," i.e., specific vocabulary to describe attributes. This offers the opportunity to more easily compare two competing technologies, which, in turn, may be compared to the Navy warfighter requirements. This effort can allow decision-makers to become aware of what programs, systems, and specific features are available for acquisition and how well they match warfighter's needs and requirements with greater effect and immediacy—possibly in real-time. We present a data-driven automation method, namely, Lexical Link Analysis (LLA), to facilitate and automate acquisition system s*elf-awareness*.

Introduction

DoD acquisition is an extremely complex system, comprised of myriad stakeholders, processes, people, activities, and organizations in an effort to provide the most useful capabilities to warfighters at the best possible value to the government. According to the *Chairman of the Joint Chiefs of Staff Instruction for Joint Capabilities Integration and Development System (JCIDS) (J-8 CJCSI 3170.01G)* (JCIDS, 2009), there are three key processes in the DoD that must work in concert to deliver the capabilities required by the warfighter: the requirements process; the acquisition process; and the Planning, Programming, Budget, and Execution (PPBE) process. In particular, the requirements process is implemented in a process called Joint Capabilities Integration and Development System (JCIDS), as shown in Figure 1. JCIDS plays a key role in identifying the capabilities required by the warfighters to support the National Defense Strategy, the National Military Strategy, and the National Strategy for Homeland Defense. The Defense Acquisition System (DAS) looks on enterprise asset acquisition based on JCIDS requirements, and PPBE is focused on the management of financial resources in accomplishing enterprise asset creation, sustainment and reuse. The leadership and decision-makers constantly contend with two major questions:

1. Are we responding to strategic guidance and joint capability needs?

    1. Are we getting the best value for taxpayers?

As shown in Figure 1, JCIDS alone produces a large amount of detailed documents (e.g., Initial Capabilities Document (ICD), Formal Capability Development Document (CDD),

for material solutions or doctrine, organization, training, materiel, leadership and education, personnel, or facilities (DOTMLPF), Change Recommendations (DCR) for non-material solutions, and Capability Production Document (CPD)). Each involves diversified stakeholders such as sponsors, program managers, developers, the Joint Requirements Oversight Council (JROC) and the Milestone Decision Authority (MDA).

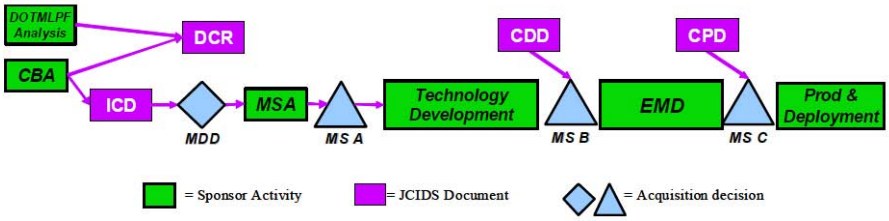


Figure 1. JCIDS Process and Acquisition Decisions
(JCIDS, 2009)

Warfighters' requirements are documented in Universal Joint Task List (UJTLs) or Joint Capability Areas (JCAs), which are collections of required capabilities functionally grouped to support mission analysis, capability analysis, strategy development, investment decision-making, capability portfolio management, and capabilities-based force development and operational planning.

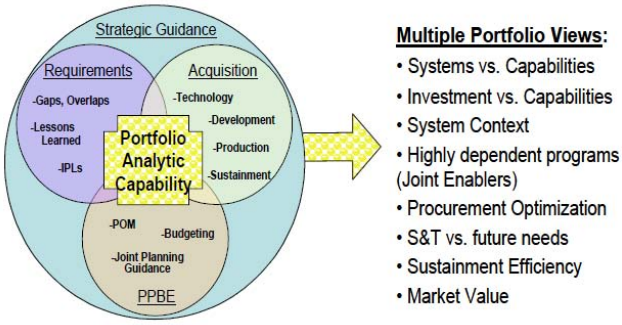


Figure 2. Portfolio Analytic Capability
(Appleton, 2009)

In summary, the major challenges in the current process can be summarized as follows:

2. To make optimal investment decisions, acquisition managers must analyze a full spectrum of data, including data that encompasses capability requirements, planning, development, integration, testing, architecture, standards, cost and schedules. This can be a daunting, if not impossible, task.

    2. The pace of technology change also requires agile decision-making and challenges program management to maintain constant awareness of what is available for acquisition.

    3. When considering an overall demand and supply in the trade space management of the Department of Defense, as shown in Figure 2, decision-makers require advanced portfolio analytic capability that can

intercept all three business processes of requirements, acquisition and PPBE under the DoD warfighting strategic guidance in the contexts of many factors, such as systems versus capabilities, investment versus capabilities, highly dependent programs, etc., in order to maximize Return of Management (ROM) and Yield on Cost (YOC) (Appleton, 2009).

4. The information produced in the process is too voluminous and unformatted to lend itself to analysis on a large scale. Decision-makers require large-scale automation and discovery tools that can speed up the analysis quickly in response to the pace of technology change, therefore adapting DoD program development and associated funding mechanisms in an agile manner. The decision-makers also require a much more fine-grained level of analysis for program-to-program and program-to-program elements analysis using the unstructured documents directly. This is a big leap that is not provided by the current analysis capabilities.

One method to reduce unknown performance measures is through participation in annual large-scale field experimentation exercises as part of the Research, Development, Test & Evaluation (RDT&E). These experiments can provide close interaction among users, developers, the test community, and decision-makers. At Distributed Information Systems Experimentation (DISE) laboratory at NPS, we collect and analyze data, help the Navy learn and manage information and knowledge resulting from large-scale annual experimentation (e.g., Trident Warrior and Empire Challenge). We believe this experiential data, together with Lexical Link Analysis methods, will produce deepened awareness of current program effectiveness for acquisition decision-makers.

Methods

Program Self-awareness

Here we consider that the cognitive interface between decision-makers and a complex system may be expressed in a range of terms or "features," i.e., specific vocabulary or lexicon, to describe attributes and the surrounding environment of a system. This process is similar or can be modeled using human cognitive processes, where the simplest form of such a model is relationships between noun/verb. In math, the model becomes variable/function; in engineering it becomes operand/operator; in information technology, it becomes data/process or description/procedure. We have borrowed from notions of "awareness," and implement the term self-awareness of a complex system as the collective and integrated understanding of system features. A related term, "situational awareness" is used in military operations and carries with it a sense of immediacy and cognitive understanding of the warfighting situation. Here, system self-awareness, or program awareness (Gallup, MacKinnon, Zhao, Robey & Odel, 2009), allows decision-makers to be aware of what systems, programs, and products are available for acquisition, how they match warfighters' needs and requirements, recognize relationships among them, improve efficiency of available collaboration, reduce duplication of effort, and re-use components to support cost effective management—with greater immediacy, possibly in real-time.

Through our research, we present a data-driven automation method, namely, a Lexical Link Analysis (LLA) for program self-awareness. This methodology is demonstrated

by extracting realistic sample data related to systems and programs included in experimentation programs, Urgent Needs Statements (UNS), and CENTCOM/NAVCENT warfighting gap/priority lists, a large-scale data set from OSD with regards to Major Defense Acquisition Programs (DMAP) and Acquisition Category II (ACATII) weapon systems and their RDT&E documentations.

Lexical Link Analysis (LLA)

Data mining includes analytic tools that may be applied to both structured and unstructured data to confirm previously determined patterns, or to discover new patterns that are yet unknown. Text mining is the application of data mining to unstructured or less structured text files. Text mining represents an emerging field with a wide range of software implementing innovative visualization and navigation techniques. These techniques graphically represent networks of documentation that are related conceptually. Visualization of relationships enables concept discovery, automated classification, and understandable categorization of unstructured documents.

Lexical Analysis (LA, 2010) is a form of text mining in which word meanings are developed from the context from which they are derived. Lexical Analysis (LA) can also be used in a learning mode, where such words and context associations are initially unknown and are constantly being "learned," updated, and improved as more data become available. Link analysis, a subset of network analysis that explores associations between objects, reveals the crucial relationships between objects when collected data may not be complete. Lexical Link Analysis (LLA) is an extended lexical analysis and link analysis enabled in a learning mode.

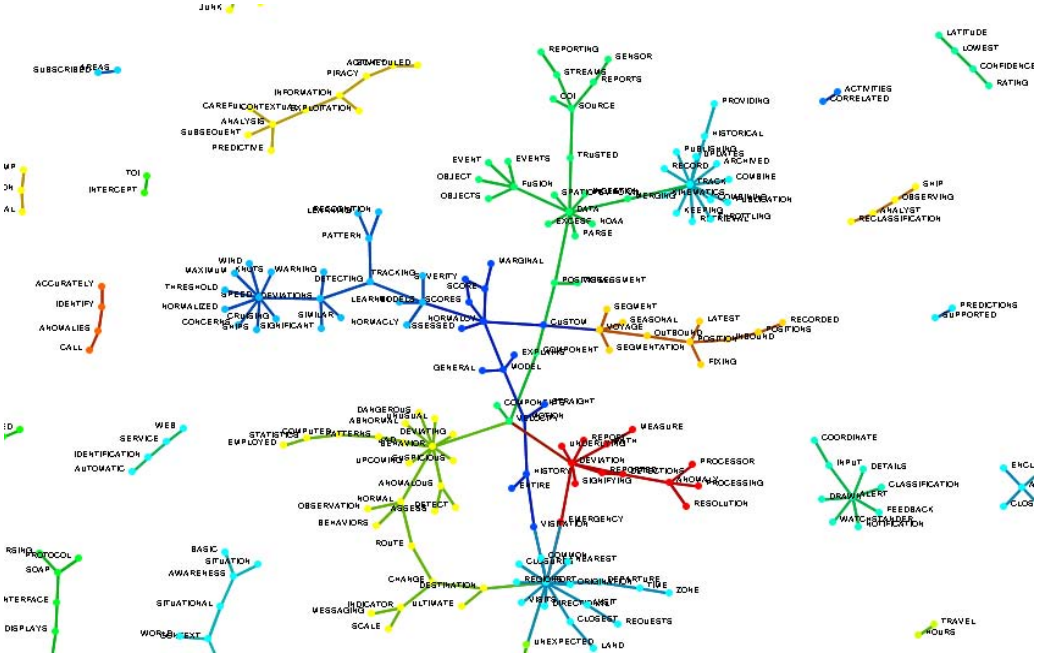


Figure 3. A Word Hub Showing the Detail on the Linkage in Figure 3

This approach clusters words and then correlates words with their textual contexts (co-occurrence), and produces a data-driven and dynamic word network. This approach is related to a number of extant tools for text mining, including Latent Semantic Analysis (LSA) (Dumais et al., 1998), advanced search engine (Foltz, 2002), key word analysis and tagging technology (Gerber, 2005), and intelligence analysis ontology for cognitive assistants (Tecuci et al., 2007). What results from this process is a learning model—like an ethnographic *code book* (Schensul, Schensul & LeCompte, 1999)—containing descriptions of both patterns and anomalies, generated using encountered terms. As an example shown in Figures 3 and 4, we applied our approach to Maritime Domain Awareness (MDA) technologies that were evaluated in Trident Warrior 08. Figure 3 shows a visualization of LLA with connected keywords or concepts extracted from the documents of MDA technologies. Words are linked as word pairs that appear next to each other in the original documents. Different colors indicate different clusters of centralization among word groups. They are produced using a link analysis method, a social network grouping method (Girvan & Newman, 2001): words are connected as shown in one color as if they are in a social community. A "hub" is a word centered with a list of other words ("fan-out" words) centered around other words. For instance, in Figure 4, the word "behavior" is centered with "suspicious, bad, dangerous, abnormal, usual, and anomalous," etc., showing the ways to describe "behavior" in the MDA area.

Figures 5 and 6 show a visualization of lexical links between Systems 1 and 2. Each node is a feature, or word hub; each color refers to the collection of lexicon (features) to describe a system, the overlapping area nodes refer to *lexical links* between systems. The nodes toward the two ends of the links represent the unique features related to each system.

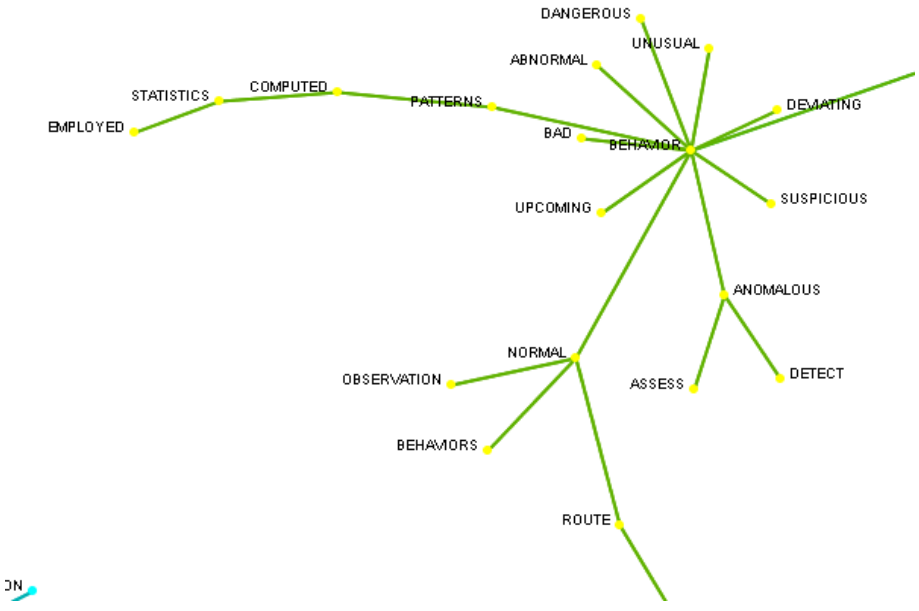


Figure 4. A Word Hub Showing the Detail on the Linkage in Figure 3

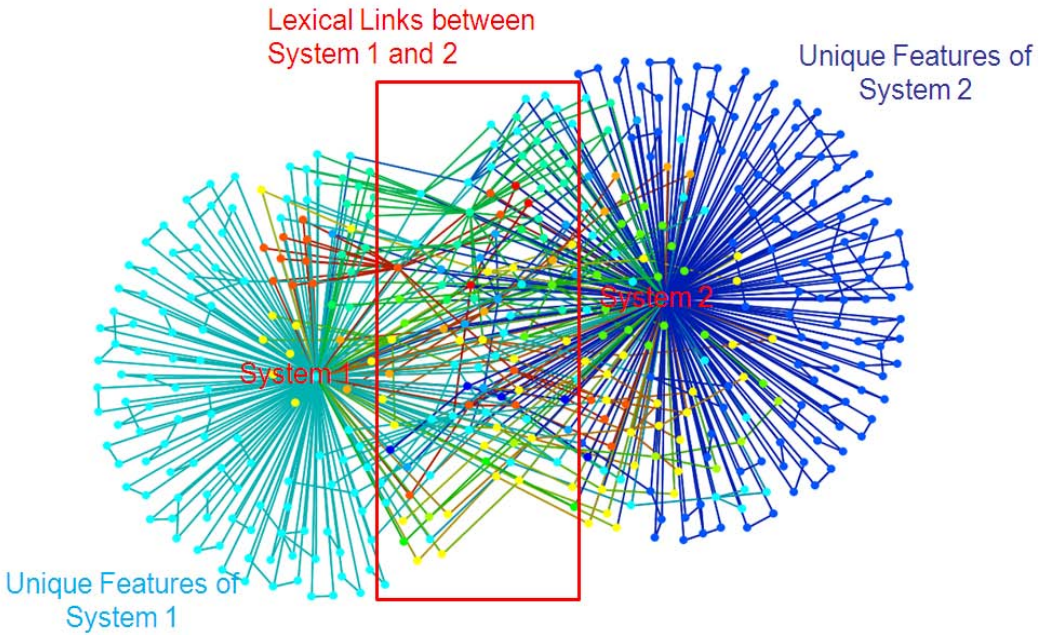


Figure 5. Visualization of Lexical Links

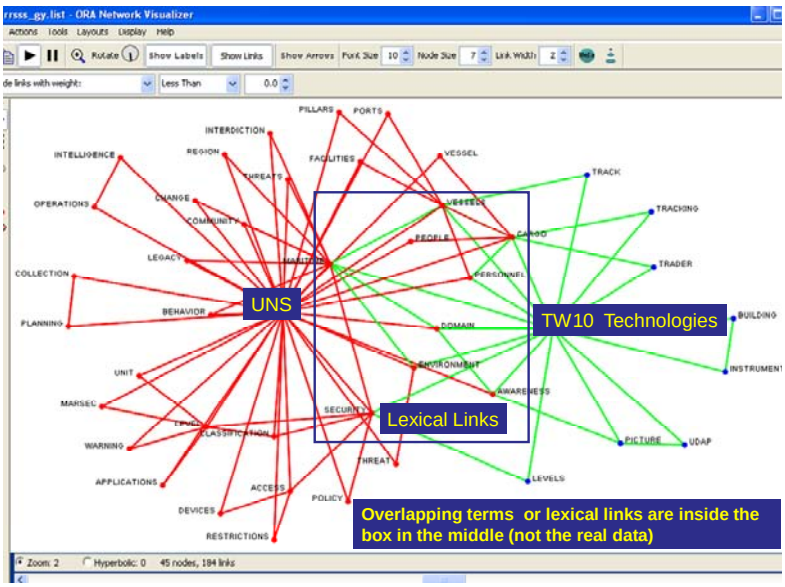


Figure 6. Overlapping Terms or Lexical Links, Shown in the Middle of Two Word Networks as the Result of the LLA Analysis

In summary, LLA provides a methodology and tools to address the following specific areas that can impact acquisition decision-making:

- LLA provides a metric to link warfighters' needs with the capabilities by directly comparing the documents that resulted from the business process—for example, linking "programs," specifically MDAPs, to operational capabilities. The number of lexical links, extracted to reflect the meaning of the documents between two systems or programs, can be a measure of consensus or synergy between the

two. This compelling perspective is central to the notion of portfolio management, for example, to answer the questions: What are the programs (e.g., MDAPs) related to a given capability? What are the gaps of warfighter requirements not addressed by current programs? Currently, human analysts are responsible to answer these questions manually. Automation is needed to facilitate human analysis and to process large volumes of data quickly.

- LLA visualization is also important for acquisition decision-making. Producing a picture illustrating where the needs are met and where the overlapping efforts and gaps are will allow decision-makers to become aware of the overall situation, thus allowing them to see trends in a larger, broader scale and in a longer timeframe. For example, combining the analyses of the Army, Navy, and Air Force from RDT&E and procurement documents might show the linkages within and among programs, as they mature from development to production. Modified programs can be illustrated to show the trend toward (or deviation away from) warfighters' needs during the program's life span. One may also visually see the resource sharing (or wasting) practices and note opportunities for growth when all the data can be summarized in a discernable picture.
- LLA discovers latent, implicit, or second-order relationships by examining the detailed budget justification documents. In general, programs retain their identities from development to production, yet may change their names or be re-designated, resulting from a milestone decision or other action. The "New Attack Sub" or "NSSN" during development, for instance, was referred to as the "Virginia Class Sub" in production. The "Joint Strike Fighter" and "F-35" are also synonymous. The official "decoder" for these transformations is the DAMIR system. We note that the mapping of MDAPs to their predecessors, successors, constituents, or dependent partners is non-trivial and is, in fact, one of the fundamental challenges for acquisition analysts.
- LLA could affect the fundamentals of acquisition processes through automation and discovery. In the defense acquisition community, decision-makers are interested in determining the costs of these programs relative to their predicted baselines (e.g., Milestone B or C). They must also determine why costs change over time. Historically, acquisition researchers only considered endogenous factors (e.g., poor program management skills) as drivers of cost changes. The notion of interdependence as a potential driver of cost may be determined by LLA. It may also help determine whether this interdependence among programs may be manifested in the sharing of resources among programs, as described by the budget artifacts. Budget artifact data are voluminous, and unstructured, which make empirical analysis extremely difficult—if not humanly impractical. Previous research has been done in this area using manually identified program interdependencies (M. Brown, personal communication, 2010) and has made great progress in establishing that interdependence exists and how they might be correlated with the program costs. LLA could automate this process of identifying interdependencies and, thus, reveal aspects of interdependence that would otherwise remain obscure.

The LLA Analysis

We began at the Naval Postgraduate School (NPS) by using Collaborative Learning Agents (CLA) (QI, 2009) and expanded to other tools, including AutoMap (AutoMap, 2009) for improved visualizations. Results from these efforts arose from leveraging intelligent agent technology via an educational license with Quantum Intelligence, Inc. CLA is a computer-based learning agent or agent collaboration, capable of ingesting and processing data sources. Each CLA is capable of revealing patterns that occur frequently and anomalies that occur rarely. Anomalies that might be interesting are thus revealed so that human analysts are alerted and can further investigate them. The CLA is able to separate the patterns from anomalies using the "patterns and anomalies separation" algorithm in each CLA to select feature-like word pairs for the LLA method.

The following are the steps for the LLA analysis:

3. Read two documents into the CLA (e.g., Urgent Needs Statement (UNS)) and a targeted technology document set (e.g., Trident Warrior 2010 (TW10).

    5. Select feature-like word pairs based on clusters using the CLA anomaly search method (Zhao & Zhou, 2008).

    6. Apply social network algorithm to group the word pairs into word categories.

    7. Apply AutoMap to visualize the associations of the requirement document set (UNS) and targeted technologies (TW10) document sets, as shown in Figures 5 and 6.

    8. Generate lexical link matrices used for further analyses, as shown in Figures 8, 9, and 10.

When mining text data or performing lexical analysis, we also apply entity extraction, known as Named Entity Recognition (NER), (NER, 2010; Nadeau, Turney & Matwin, 2006), which recognizes named entities such as persons, organizations, locations, expressions of times, quantities, monetary values and percentages in context. The extracted entities could also be examined separately. Excluding these modifiers from the terms resulting from Lexical Link Analysis (LLA) can provide an improved comparison by focusing on term semantics.

In some applications, differentiating nouns from verbs and adjectives, or having the ability to parse the syntax into nouns, verbs, subjects, and objects, could be helpful to acquisition managers to develop understanding. We also use a Part-of-Speech (POS) *tagger* as pre- or post-processing filters for this purpose. A POS tagger is a piece of software that reads text in some language and assigns parts of speech to each word, such as a noun, verb, adjective, etc. We have chosen the Stanford Natural Language Processing (NLP) tool (Toutanova, Klein, Manning & Singer, 2003; Stanford NLP, 2009) to perform this task. The POS taggers are usually language dependent. Our method is statistically based and can, therefore, employ NER and POS as pre- or post-processing filters.

Data Sets

We report a case study using LLA comparing US Navy Urgent Need Statements (UNS) with Trident Warrior 10 Technologies. The goal is to compare the two respective data sets, the first one is an Excel file (UNS.xls) representing Urgent Need Statements collected from C4I users. Each urgent need is listed as a statement. The UNS.xls is classified; therefore, details of this document set are not reported in this paper. The second data set is called "Focus Area Assignment TW 10.xls," also in an Excel format. It includes information from each selected technology in Trident Warrior 10.

Trident Warrior (TW) is an annual Navy FORCEnet operational experiment. At the Distributed Information Systems Experimentation (DISE) laboratory at NPS, we collect and analyze data from this and other experimentation venues to help the Navy learn and manage information and knowledge resulting from large field experiments such as Trident Warrior to provide a basis for DoD acquisition of systems and technologies. The technology information includes each technology's objective(s) for the experimentation, including Concept of Operations (e.g., how a warfighter will utilize it), and what each technology provider intends to learn from the experimentation (e.g., decrease timeline, standardized process, and/or reduced workload, etc.). TW data also includes decisions that may affect experimentation findings.

Result Presentation and Visualization Tools

Figure 7 illustrates a result summary revealing terms or word pairs combined into word categories, displayed in a radial graph. The categories with radius = 2 represent overlapping word categories that are found in both requirements (UNS) and technologies (TW10). The categories with radius = 1 indicate where gaps exist, i.e., terms that show in the UNS but not in the TW10 technologies or vice versa. We determine that there is between a 60% and 70% match overlap of technology correlations between UNS and TW 10 technologies. For example, 42 of 67 (62%) of the UNS word categories matched (were served by) with TW10 technologies.

In addition, word network views of lexical links are produced using a network tool, AutoMap. We also developed several outputs to view the detailed LLA analysis results as shown in Figures 8, 9, and 10. Figure 8 shows an Excel document output, including a few columns of information as follows:

- Terms: Matching terms or word categories discovered automatically via the LLA method.
- UNS: Values can be 0, 1, 2, specifically:
  - 0: terms not found in UNS,
  - 1: terms only found in UNS, and
  - 2: terms found in both UNS and TW10.
- UNS IDS: UNS documents in which the terms can be found.
- TW10: Values can be 0, 1, 2.
  - 0: terms not found in TW10,
  - 1: terms only found in UNS, and

- 2: terms found in both UNS and TW10.
- TW10 IDS: TW10 documents in which the terms can be found.
- Tech Features: Terms only belong to TW10.

- As one scrolls down, if there is “0” in the TW10 column, then it indicates a gap area for TW10. Similarly, in scrolling further, if there is a “0” in the UNS column, then this indicates a gap in UNS.

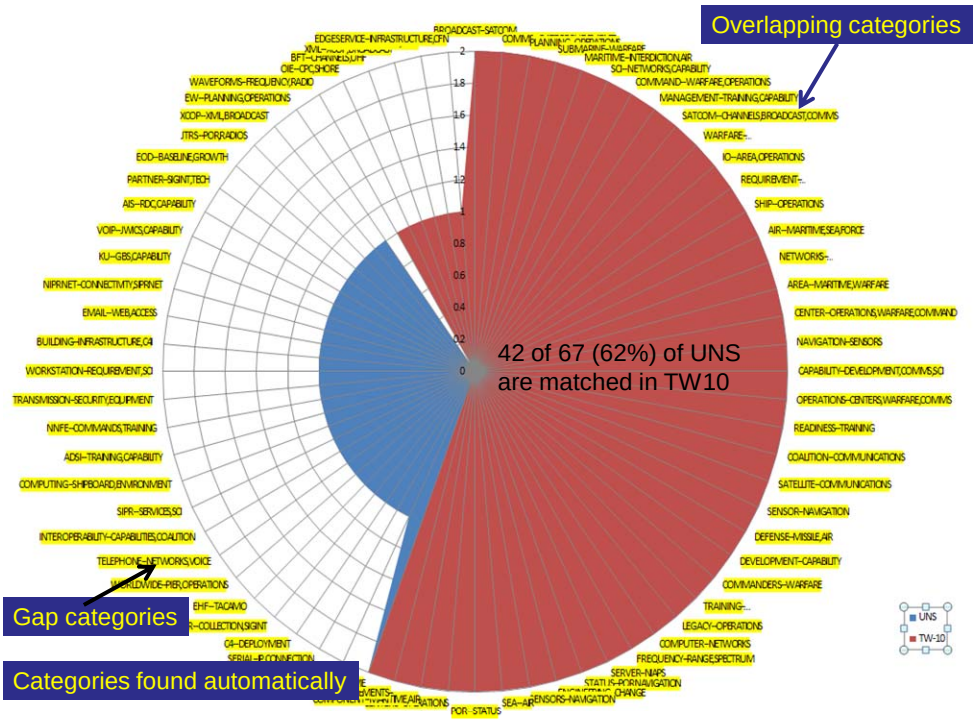


Figure 7. Resulting View and Visualization Illustrating “Overlapping” and “Gap” Word Categories

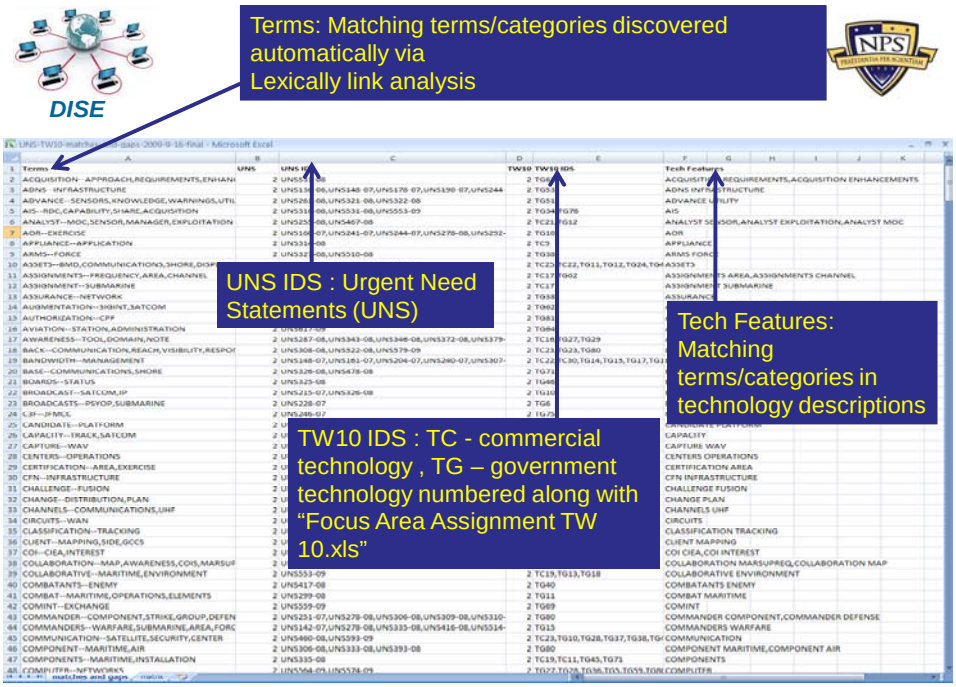


Figure 8. The Spreadsheet View of the LLA Analysis with "Matched" Terms and "Gap" Terms

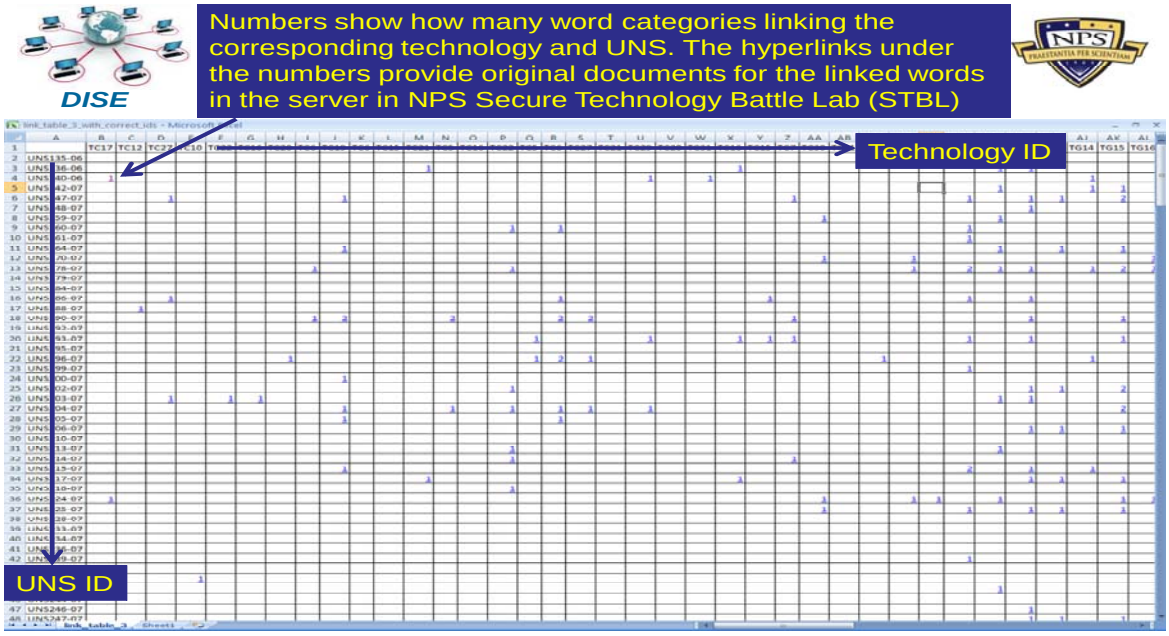


Figure 9. The Matrix View of the LLA Analysis

Figure 9 shows a matrix view of UNS to TW 10 technologies. Where numbers are seen indicates a numerical reference to the number of the "concepts" (terms or word categories) included between UNS and technologies that are being satisfied. Usually, there are multiple concepts within a UNS statement and a tech description. Each number is also a hyperlink back to the original document in a server where it is stored, e.g., the server in the NPS Secure Technology Battle Lab (STBL) for classified documents.

These results can be increasingly focused as the Intelligent Agent (IA) becomes "tuned," or *learns* what it is that the researcher is attempting to understand. This effort can then become increasingly automated.

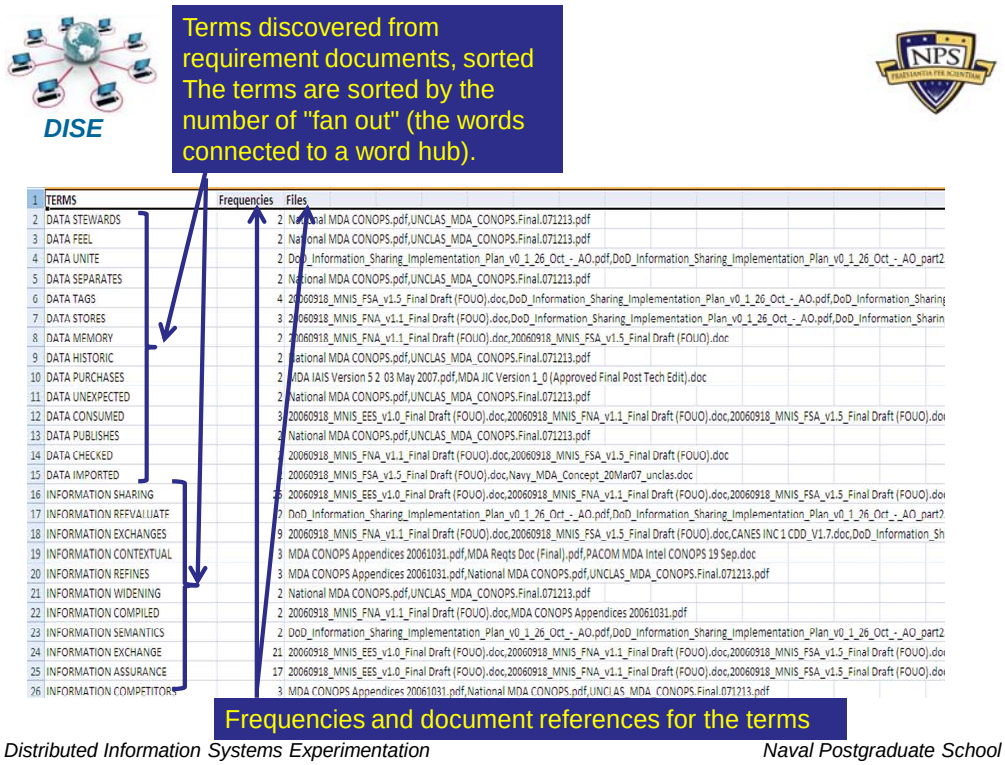


Figure 10. Frequency Count and Document References

Figure 10 shows a summary spreadsheet listing the terms and number of files in which the terms appear. This output can be used to discover concepts (terms) that are cross-validated by at least two documents in a document set. The terms are sorted by the number of "fan out" (the words connected to a word hub), showing the critical concepts being addressed across multiple documents. The top few sorted word groups, e.g., “data” and “information” in this case, are the key requirements that result in substantial consensus across different levels of requirement generation mechanisms—for example, Joint Integrating Concept (JIC), Joint Capability Areas (JCA), the Universal Joint Task List (UJTL), and user communities such as US Northern Command, US Pacific Command, and sponsors that are interested in Interagency Investment Strategies (IISs).

Validity

Several methods are being investigated to validate LLA methods. Currently, we have shown these proof-of-concept results to Subject-matter Experts (SME) from various organizations (e.g., Joint Force Development and Integration, the J-7 Staff) for evaluation and comment. One MDA expert has commented on the summary spreadsheet by saying, “it is very useful, particularly the frequency count and the documented reference.” Other SMEs comment that “LLA has great potential to help us link the UNS with the technology and further fill in the gaps that are out there.” “This would be highly useful and has great potential to help us in the larger N9/Sea Trial construct and spoke further of the possibility of using LLA at the Joint Warfighter Challenges level.” We will consider quantitative content validation methods between SMEs and LLA, such as correlation and inter-rater reliability scores (Cohen's Kappa; Kerlinger & Lee, 1992), as well as large-scale correlation

calculation used in sections below.

Towards a Large-Scale Example of Program Self-Awareness

We have worked with OUSD(AT&L)/ARA/EI on the broader data sets and a large-scale application of program self-awareness via LLA.

Data Sets

Figure 11. DoD Budget Documentation

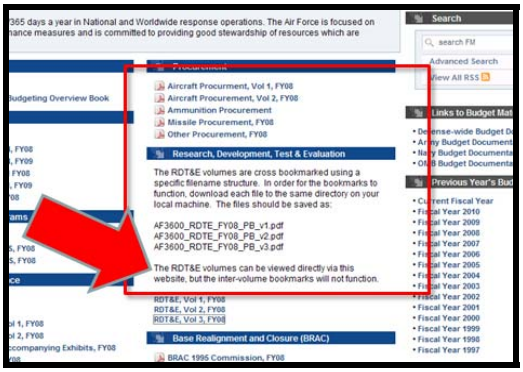


Figure 12. Research, Development Test & Evaluation (RDT&E)

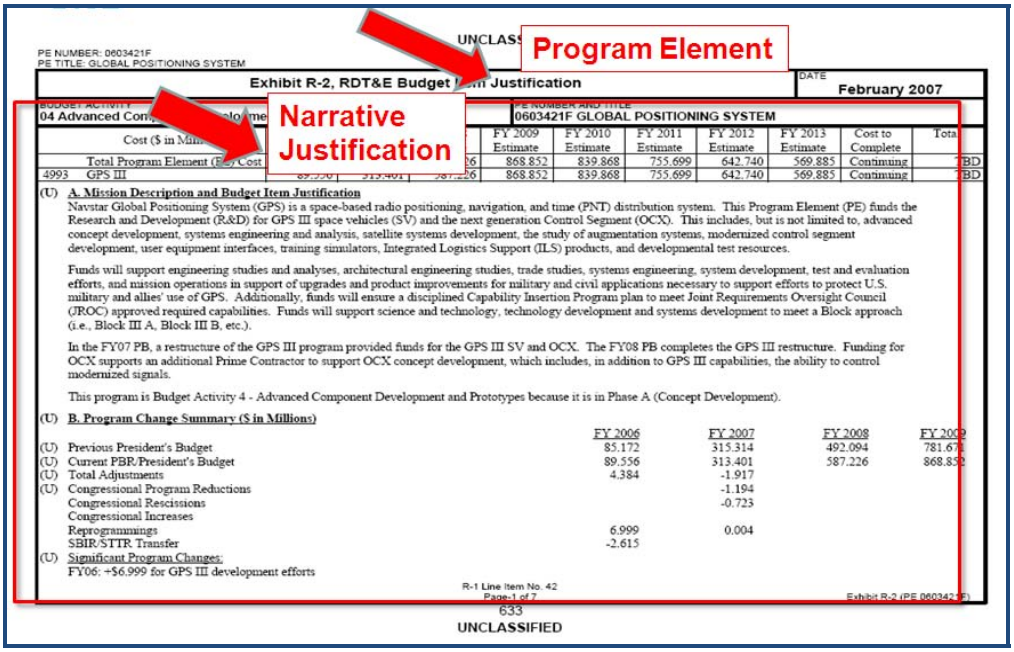


Figure 13. Program Element RDT&E Budget Justification

4. We have obtained program element (PE) data, which are used for DoD budget justification each year, as shown in Figure 11. One PE component is Research, Development, Test & Evaluation, which is the budget estimation, allocation and justification used for programs in the earlier stages of

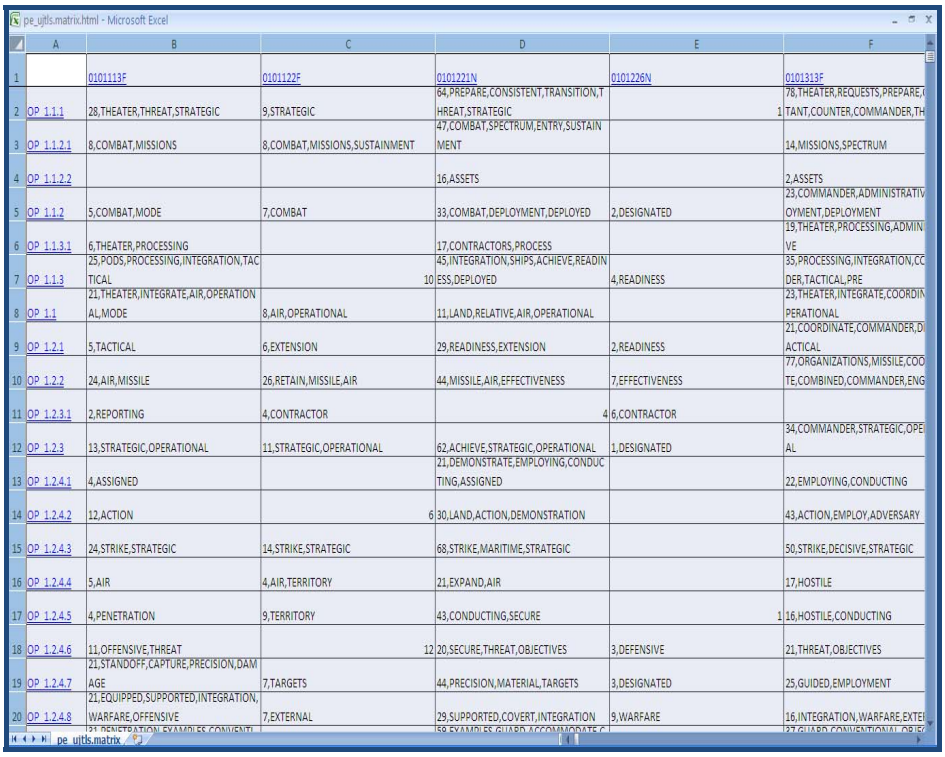
development. The procurement of PE components is the counterpart used for mature products. RDT&E books are obtained from the Air Force, Army (http://asafm.army.mil/Document.aspx?OfficeCode=1200) and Navy (http://www.finance.hq.navy.mil/fmb/11pres/BOOKS.htm) websites.

9. The *Weapon Book* (Weapon, 2008), which summarizes weapons and their basic functions and missions, combined total cost from RDT& and procurement.
10. MMT databases contain cost and schedule information for each program. They consist of MDAPs and weapon systems. MMT databases also contain various program interdependencies identified by human analysts that can be used for validation. MMT databases also contain JCAs and UJTLs mapped to programs that are handmade by human experts.

According to program managers Data (1) and (2) are so voluminous, unformatted and unstructured that traditional analysis methods are difficult to apply on this scale; therefore, they are the major focuses of the analysis for LLA. There are about ~500 PEs and ~80 weapon systems extracted from data sets (1) and (2), with a total size about ~200M. Data (3) is unstructured and various previous research has been conducted on this data and, therefore, can be used to validate the LLA method against human analyses.

LLA Analysis

The focus of this paper is to show that the LLA method is capable of improving system self-awareness. LLA is able to produce this by providing an improved methodology and toolset for automation and discovery of patterns and anomalies within structured and unstructured data. This discovery can be used to produce graphics illustrating gaps and overlaps existing between systems and the needs of the DoD by basing comparisons on the *features* of each system. This methodology can have the effect of improved savings for the DoD, while developing high-value products that meet warfighters' needs.



	A	B	C	D	E	F
1		0101113F	0101122F	0101221N	0101226N	0101313F
2	OP 1.1.1	28,THEATER,THREAT,STRATEGIC	9,STRATEGIC	64,PREPARE,CONSISTENT,TRANSITION,THREAT,STRATEGIC	1	78,THEATER,REQUESTS,PREPARE,TANT,COUNTER,COMMANDER,TH
3	OP 1.1.2.1	8,COMBAT,MISSIONS	8,COMBAT,MISSIONS,SUSTAINMENT	47,COMBAT,SPECTRUM,ENTRY,SUSTAINMENT		14,MISSIONS,SPECTRUM
4	OP 1.1.2.2			16,ASSETS		2,ASSETS
5	OP 1.1.2	5,COMBAT,MODE	7,COMBAT	33,COMBAT,DEPLOYMENT,DEPLOYED	2,DESIGNATED	23,COMMANDER,ADMINISTRATIVOYMENT,DEPLOYMENT
6	OP 1.1.3.1	6,THEATER,PROCESSING		17,CONTRACTORS,PROCESS		19,THEATER,PROCESSING,ADMINVE
7	OP 1.1.3	25,PODS,PROCESSING,INTEGRATION,TACTICAL	10	45,INTEGRATION,SHIPS,ACHIEVE,READINESS,DEPLOYED	4,READINESS	35,PROCESSING,INTEGRATION,CCDER,TACTICAL,PRE
8	OP 1.1	21,THEATER,INTEGRATE,AIR,OPERATIONAL,MODE	8,AIR,OPERATIONAL	11,LAND,RELATIVE,AIR,OPERATIONAL		23,THEATER,INTEGRATE,COORDINPERATIONAL
9	OP 1.2.1	5,TACTICAL	6,EXTENSION	29,READINESS,EXTENSION	2,READINESS	21,COORDINATE,COMMANDER,DACTICAL
10	OP 1.2.2	24,AIR,MISSILE	26,RETAIN,MISSILE,AIR	44,MISSILE,AIR,EFFECTIVENESS	7,EFFECTIVENESS	77,ORGANIZATIONS,MISSILE,COOTE,COMBINED,COMMANDER,ENG
11	OP 1.2.3.1	2,REPORTING	4,CONTRACTOR	4	6,CONTRACTOR	
12	OP 1.2.3	13,STRATEGIC,OPERATIONAL	11,STRATEGIC,OPERATIONAL	62,ACHIEVE,STRATEGIC,OPERATIONAL	1,DESIGNATED	34,COMMANDER,STRATEGIC,OPEAL
13	OP 1.2.4.1	4,ASSIGNED		21,DEMONSTRATE,EMPLOYING,CONDUCTING,ASSIGNED		22,EMPLOYING,CONDUCTING
14	OP 1.2.4.2	12,ACTION	6	30,LAND,ACTION,DEMONSTRATION		43,ACTION,EMPLOY,ADVERSARY
15	OP 1.2.4.3	24,STRIKE,STRATEGIC	14,STRIKE,STRATEGIC	68,STRIKE,MARITIME,STRATEGIC		50,STRIKE,DECISIVE,STRATEGIC
16	OP 1.2.4.4	5,AIR	4,AIR,TERRITORY	21,EXPAND,AIR		17,HOSTILE
17	OP 1.2.4.5	4,PENETRATION	9,TERRITORY	43,CONDUCTING,SECURE	1	16,HOSTILE,CONDUCTING
18	OP 1.2.4.6	11,OFFENSIVE,THREAT	12	20,SECURE,THREAT,OBJECTIVES	3,DEFENSIVE	21,THREAT,OBJECTIVES
19	OP 1.2.4.7	21,STANDOFF,CAPTURE,PRECISION,DAMAGE	7,TARGETS	44,PRECISION,MATERIAL,TARGETS	3,DESIGNATED	25,GUIDED,EMPLOYMENT
20	OP 1.2.4.8	21,EQUIPPED,SUPPORTED,INTEGRATION,WARFARE,OFFENSIVE	7,EXTERNAL	29,SUPPORTED,COVERT,INTEGRATION	9,WARFARE	16,INTEGRATION,WARFARE,EXTE

Figure 14. An Example of LLA Matrices of Program Elements (PE) against UJTLs

First, we want to show how LLA provides a new metric to measure how warfighters' needs are matched with resources and products that are being considered. Figure 14 shows an LLA matrix result using program elements as columns and UJTLs as rows. The number in each cell is a match score generated from the LLA method. Next to the score are word hubs that indicate which term is matched. Sorting this matrix according to the matched scores vertically and horizontally answers the following questions:

- Which programs (e.g., MDAPS) are related to a given capability? Which PEs are related to a given capability?
- How is the acquisition process responding to expressed capability needs? How much of the weapon systems acquisition budget is being allocated to any given operational need (e.g., UJTL).

Note that this LLA matrix can be generated for any pair of document collections that are desired for comparison, e.g., PEs versus UJTLs, weapon systems versus UJTLs and weapon systems versus weapon systems. When applied to weapon systems (MDAPs) versus UJTLs, we can answer the following question by sorting the LLA matching scores:

- Which capability(ies) does any given MDAP support? How much does the MDAP contribute to this capability?

The LLA matrices may also help to reconcile the gaps between the final products and what warfighters need after the long process of design and development. Furthermore, they may also provide new prospective for portfolio analysis. A conventional treatment of portfolio analysis is that it is typically expressed as a simple correlation between an MDAP and a capability. This simple correlation ignores the fact that no individual program (system,

platform, etc.) can contribute to any capability unless other programs/systems/capabilities are in place. The analogy is that a fighter jet is useless unless it has all the supporting capabilities/infrastructure (airfield, ammo, fuel, personnel, etc.), and complementary systems (e.g., GPS, C2, satellite imagery, mission planning, etc.) to enable it to operate effectively. Considering a single MDAP in terms of how much it contributes to a given capability without considering its linkages to other systems/programs/capabilities might be counterproductive, and would likely drive bad decisions. The better approach is to consider a program in the context of its interdependencies with respect to their collective contribution to a specific capability. The interdependencies should be identified from operational needs, engineering constructions and programmatic budget justifications. Therefore, the combinations of the LLA matrices—for example, PEs versus UJTLs, weapon systems versus UJTLs and weapon systems versus weapon systems may also help to redefine portfolios and improve portfolio management.

Validity

In order to realize the potential of the LLA method, an important first step is to establish the validity of the method in the context of realistic large-scale data sets. For that, we used the matrix generated from PEs versus PEs, compared with what human analysts have identified previously. As shown in Figure 15, in each program element artifact, another program element might be referenced, indicted as precedent or directionally linked program elements. A *backward* link is usually a stronger indicator of importance of a PE than a *forward* link. This is similar to the information retrieval or page ranking in a search engine (e.g., Google). Here, we use the number total *forward* and *backward* links together, identified by human analysts, as the attributes to validate the LLA method. For example, Figure 15, PE 0604602F references PE 0605011F, in which we define it as a *forward* link, for PE 0604602F; while PE 0605011F is referenced by PE 0604602F, which we define as a *backward* link, for PE 0605011F. As shown in Figure 16, the top yellow row contains the total number of unique word hubs for a PE, matched with all PEs other than itself; and the bottom yellow row contains the total number of forward and backward links for the same PE. The Pearson correlation of the two rows is 0.39, with a p-value < 0.0000001 (bi-directional t-test with a sample size N=461). This indicates that the positive correlation between the LLA-identified links and human-analyst-identified links is statistically significant and, therefore, is a validation for the LLA method.

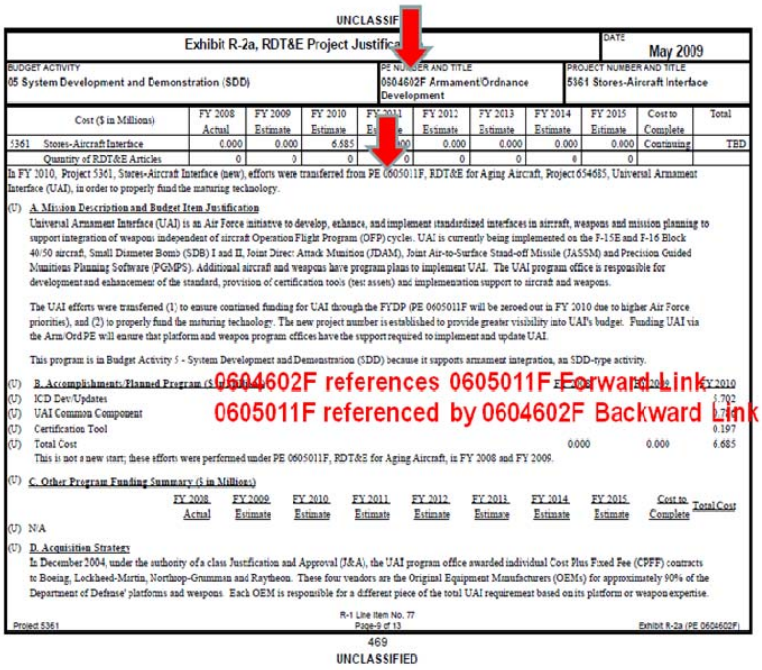


Figure 15. Program Element Cross-References Identified by Human Analysts

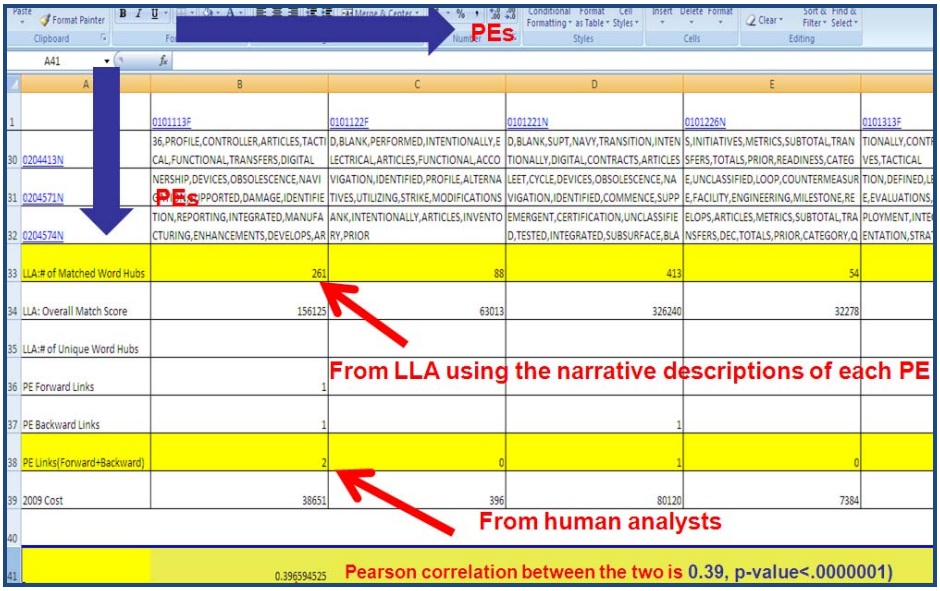


Figure 16. The Correlation Between LLA Word Hubs and PE Links Identified by SME's is Statistically Significant

Acquisition Decision-making

To support effective decision-making, we need to form a full understanding of a program in context; we need to understand the linkages and interdependencies across the operational, constructive, and programmatic domains.

An LLA matrix using programs such as weapon systems as rows as well as columns is shown in Figure 17. The lexical links output from this view show the relationships among weapon systems, therefore representing a constructive view of programs in context. The hypothesis is that more lexical links among programs may be correlated with the overall higher program total costs. The correlation between the overall LLA match score and the program total cost found in the weapon data—which includes RDT&E and procurement costs together—is 0.21, with a p-value < 0.032. This indicates there is a statistically significant relationship between the number of lexical links as an interdependency measures among programs and total cost of programs.

Similarly, a programmatic view of an LLA matrix can be generated by using weapon systems as columns and program elements as rows. The correlation between the overall LLA match scores and total program costs is 0.13 with a p-value < 0.12. This indicates that this correlation is not statistically significant based on the analyzed data.

An operational view of the LLA matrix was generated by using weapon systems as columns and UJTLs as rows. The correlation between the overall LLA match scores and total program costs is 0.086, with a p-value < 0.12, indicating that this correlation is not statistically significant.

From an acquisition management and resource analysis perspective, we conclude that

- Major programs are interdependent on one another. Interdependence can be shown by their lexical links in budget documentations in constructive, programmatic and operational views. The degree that programs are interdependent can be measured by the number of lexical links.
- Highly interconnected programs in a constructive view are statistically significantly and more expensive than less-interconnected programs (correlation 0.21, p-value < 0.032). The word hubs selected from LLA suggest the “threads” that link a portfolio of programs through shared resources. As an example, in Figure 18 ADVANCED MEDIUM RANGE AIR-TO-AIR MISSILE (AMRAAM) and AIR INTERCEPT MISSILE – 9X (AIM-9X) are connected through “COUNTERMEASURES,” which may share resources from PE 030140N.

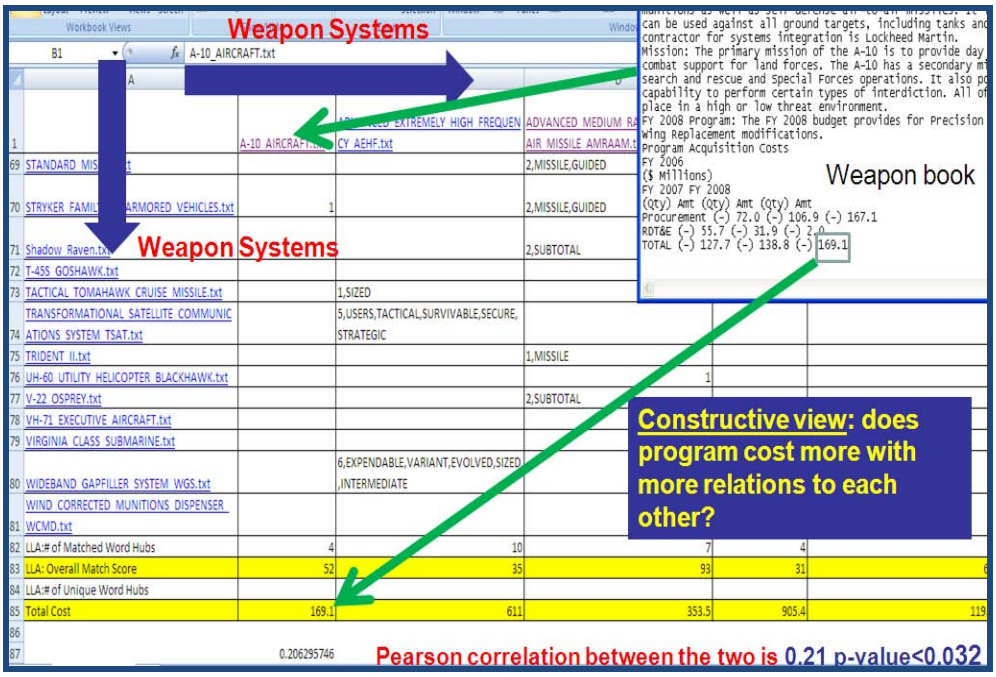


Figure 17. A Constructive View: An LLA Matrix Weapon Systems versus Weapon Systems

(Note: The correlation between the LLA overall match scores and total program costs is statistically significant.)

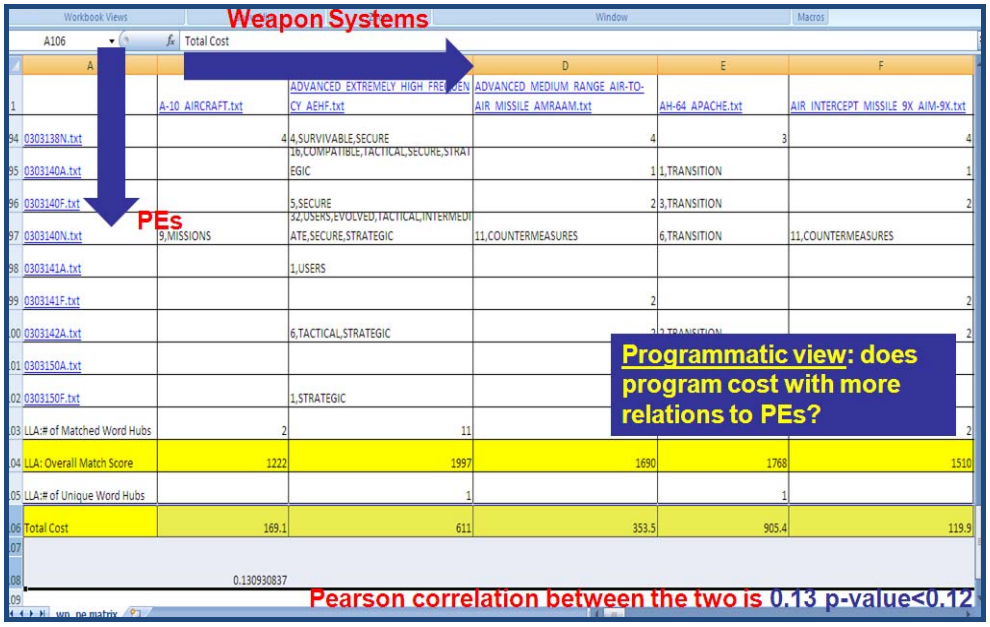
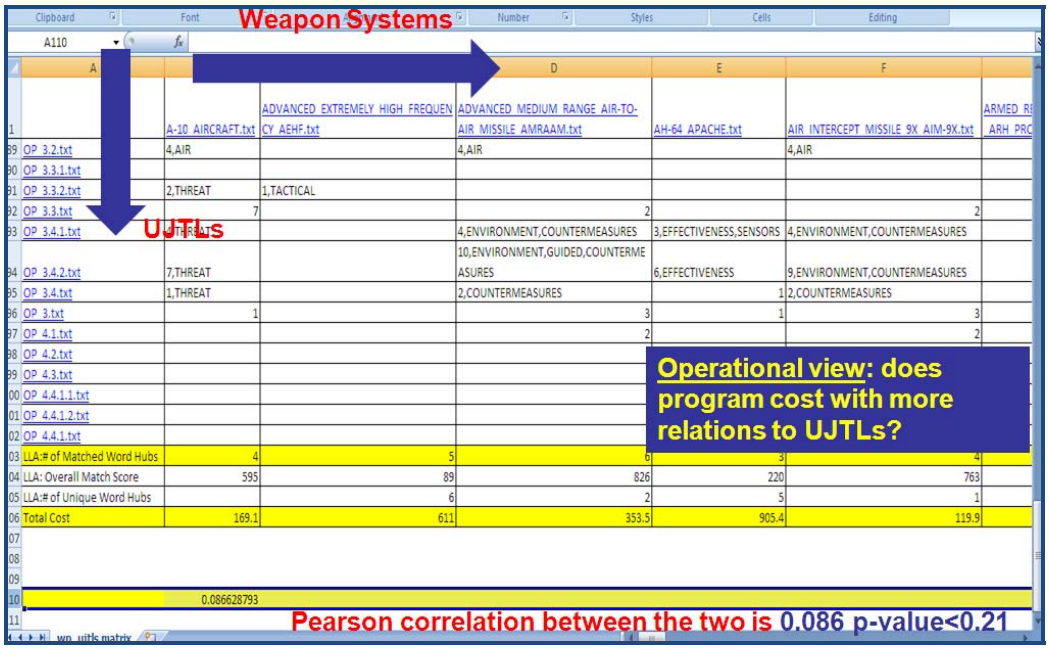


Figure 18. A Programmatic View: Weapon Systems versus Program Elements



	A	A-10 AIRCRAFT.txt	ADVANCED EXTREMELY HIGH FREQUENCY AEHF.txt	ADVANCED MEDIUM RANGE AIR-TO-AIR MISSILE AMRAAM.txt	AH-64 APACHE.txt	AIR INTERCEPT MISSILE 9X AIM-9X.txt	ARMED RE ARH PRO
89	OP 3.2.txt	4,AIR		4,AIR		4,AIR	
90	OP 3.3.1.txt						
91	OP 3.3.2.txt	2,THREAT	1,TACTICAL				
92	OP 3.3.txt	7		2		2	
93	OP 3.4.1.txt	THREAT		4,ENVIRONMENT,COUNTERMEASURES	3,EFFECTIVENESS,SENSORS	4,ENVIRONMENT,COUNTERMEASURES	
94	OP 3.4.2.txt	7,THREAT		10,ENVIRONMENT,GUIDED,COUNTERMEASURES	6,EFFECTIVENESS	9,ENVIRONMENT,COUNTERMEASURES	
95	OP 3.4.txt	1,THREAT		2,COUNTERMEASURES	1	2,COUNTERMEASURES	
96	OP 3.txt	1		3	1	3	
97	OP 4.1.txt			2		2	
98	OP 4.2.txt						
99	OP 4.3.txt						
00	OP 4.4.1.1.txt						
01	OP 4.4.1.2.txt						
02	OP 4.4.1.txt						
03	LLA:# of Matched Word Hubs	4	5	6	3	4	
04	LLA: Overall Match Score	595	89	826	220	763	
05	LLA:# of Unique Word Hubs		6	2	5	1	
06	Total Cost	169.1	611	353.5	905.4	119.9	
07							
08							
09							
10		0.086628793					
11							

Figure 19. An Operational View: Weapon Systems versus UJTLS

Our near-term plan is to apply the method jointly with the unstructured data with the MMT databases to illustrate if the LLA method can be used to address the following questions:

5. The narrative sections reference program–to-program interdependencies (e.g., Wideband Gapfiller System flies on an EELV launch vehicle). How could this be compared with program interdependence information from the DAES, or the ISP from our data set?

    11. Are these programs more or less likely to incur cost growth relative to their milestone B baselines? Are they more or less likely to breach their cost/schedule/performance baselines?

    12. How do we determine the correlation using metrics that fundamentally affect acquisition decision-making? For example, total program cost and cost growth relative to the Milestone B baseline cost. (To do that, we would need to capture the total program cost (development, procurement, and the two combined) estimated at milestone B, and compare that with these values at milestone C. These data are in the MMT data set.)

    13. Can LLA of budget documentation provide an aggregate dollar figure that describes the value/magnitude of resources being shared among these entities? Is this a reasonable proxy for the degree or significance of interdependence?

    14. Is there additional latent risk to programs that share resources? Is there potential for unanticipated "ripple effect" that could magnify budget perturbations? Can these effects be modeled or predicted? Would this suggest that new approaches to budget analysis are needed?

Large-scale and Real-time Consideration

A large number of CLA agents work together in a parallel fashion. This allows the LLA method to scale up to distributed, large-scale and real-time data sources. At the time of this printing, we have prototyped a multi-agent network of ~10 to 100 agents in the NPS High Performance Computing Center (HPC) in the Hamming Linux Cluster (HLC), which provides the requisite supercomputing for the visualization of the results. Servers are also being built in the NPS Secure Technology Battle Lab (STBL) to process classified data.

Conclusion

We show in this paper how to use the Lexical Link Analysis (LLA) to match system features with those defined in the original requirements, discover relationships among systems, and identify gaps with respect to warfighters' needs. We initially validate the LLA method and show results by correlating program interdependencies resulted from the LLA method with those from subject-matter experts. The Pearson correlation for a sample of 461 program elements (PEs) is 0.39 with a p-value < 0.0000001. This indicates the positive correlation between the LLA identified links as compared to human-analyst-identified links and that they are reasonably correlated with statistical significance. We also found that Major Defense Acquisition Programs (MDAP's) are interdependent from one another and that such interdependence can be shown by their lexical links in documentations in constructive, programmatic, and operational views. The number of lexical links can be used as a metric to measure interdependencies among new technologies. Highly interconnected programs in a constructive view are statistically significantly and more expensive than the less-interconnected programs (correlation 0.21, p-value < 0.032). Ultimately, in this vein, we seek to use the LLA method to automate and improve program self-awareness and make it feasible for acquisition decision-makers to analyze and dynamically monitor large-scale acquisition documents. The resulting system analyses will facilitate real-time program awareness and can reduce the workload of decision-makers who would otherwise perform the relations-building task manually, thus making a profound impact on the agility and perhaps the long-term success of acquisition strategies.

Acknowledgements

We thank Mr. Dave Summer from NNWC for helping us understanding warfighter requirements in the Maritime Domain Awareness (MDA) area. We thank Mr. Robert Flowe from OSD, who pointed to large-scale data sets, provided critical acquisition research, and relevant questions along with insightful discussions.

References

Appleton, D.S. (2009). *Trade space management in the Department of Defense*.

AutoMap. (2009). Retrieved from http://www.casos.cs.cmu.edu/projects/automap/

Dumais, S.T., et al. (1988). Using latent semantic analysis to improve information retrieval. In *Proceedings of CHI'88: Conference on Human Factors in Computing* (pp. 281-285).

Foltz, P.W. (2002). Quantitative cognitive models of text and discourse processing. In *The Handbook of Discourse Processes*. Mahwah, NJ: Lawrence Erlbaum Publishing.

Gallup, S.P., MacKinnon, D.J., Zhao, Y., Robey, J., & Odell, C. (2009, October 6-8). Facilitating decision making, re-use and collaboration: A knowledge management approach for system self-awareness. In *Proceedings of the International Joint Conference on Knowledge Discovery, Knowledge Engineering, and Knowledge Management (IC3K)*. Madeira, Portugal.

Gerber, C. (2005). Smart searching, new technology is helping defense intelligence analysts sort through huge volumes of data. *Military Information Technology*, *9*(9).

Girvan, M., & Newman, M.E.J. (2002). *Community structure in social and biological networks*.

JCIDS. (2009). *Chairman of the joint chiefs of staff instruction for joint capabilities integration and development system* (JCIDS) (J-8 CJCSI 3170.01G). Retrieved from http://www.intelink.sgov.gov/wiki/JCIDS

Kerlinger, F., & Lee, H. (1992). *Foundations of behavioral research* (3rd ed.). Wadsworth.

LA. (2010). Retrieved from http://en.wikipedia.org/wiki/Lexical_analysis

Nadeau, D., Turney, P.D., & Matwin, S. (2006). Unsupervised named-entity recognition: Generating gazetteers and resolving ambiguity. In *Proceedings of the Canadian Conference on Artificial Intelligence*.

NER. (2010). Retrieved from http://en.wikipedia.org/wiki/Named_entity_recognition

QI. (2009). Retrieved from http://www.quantumii.com

Schensul, J.J., Schensul, S.L., & LeCompte, M.D. (1999). *Essential ethnographic methods: Observations, interviews and questionnaires*. Rowman Altamira.

Tecuci G., Boicu, M., Marcu, D., Boicu, C., Barbulescu, M., Ayers, C., & Cammons, D. (2007). Cognitive Assistants for Analysts. In J. Auger & W. Wimbish (Eds.), Joint Publication of the National Intelligence University, Office of the Director of National Intelligence, and US Army War College Center for Strategic Leadership.

Toutanova, K., Klein, D., Manning, C., & Singer, Y. (2003). Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In *Proceedings of HLT-NAACL* (pp. 252-259).

Weapon. (2008). Retrieved from http://comptroller.defense.gov/defbudget/fy2008/fy2008_weabook.pdf

Zhao, Y., & Zhou, C. (2008a). *Fusion and visualization for multiple anomaly detection systems* (US Patent 12/042338).

Zhao, Y., & Zhou, C. (2008b). *Knowledge pattern search from networked agents* (US Patent 20080086436).

Stanford NLP. (2009). Retrieved from http://nlp.stanford.edu/software/tagger.shtml

2003 - 2010 Sponsored Research Topics

Acquisition Management

- Acquiring Combat Capability via Public-Private Partnerships (PPPs)
- BCA: Contractor vs. Organic Growth
- Defense Industry Consolidation
- EU-US Defense Industrial Relationships
- Knowledge Value Added (KVA) + Real Options (RO) Applied to Shipyard Planning Processes
- Managing the Services Supply Chain
- MOSA Contracting Implications
- Portfolio Optimization via KVA + RO
- Private Military Sector
- Software Requirements for OA
- Spiral Development
- Strategy for Defense Acquisition Research
- The Software, Hardware Asset Reuse Enterprise (SHARE) repository

Contract Management

- Commodity Sourcing Strategies
- Contracting Government Procurement Functions
- Contractors in 21st-century Combat Zone
- Joint Contingency Contracting
- Model for Optimizing Contingency Contracting, Planning and Execution
- Navy Contract Writing Guide
- Past Performance in Source Selection
- Strategic Contingency Contracting
- Transforming DoD Contract Closeout
- USAF Energy Savings Performance Contracts
- USAF IT Commodity Council
- USMC Contingency Contracting

Financial Management

- Acquisitions via Leasing: MPS case
- Budget Scoring
- Budgeting for Capabilities-based Planning

- Capital Budgeting for the DoD
- Energy Saving Contracts/DoD Mobile Assets
- Financing DoD Budget via PPPs
- Lessons from Private Sector Capital Budgeting for DoD Acquisition Budgeting Reform
- PPPs and Government Financing
- ROI of Information Warfare Systems
- Special Termination Liability in MDAPs
- Strategic Sourcing
- Transaction Cost Economics (TCE) to Improve Cost Estimates

Human Resources

- Indefinite Reenlistment
- Individual Augmentation
- Learning Management Systems
- Moral Conduct Waivers and First-tem Attrition
- Retention
- The Navy's Selective Reenlistment Bonus (SRB) Management System
- Tuition Assistance

Logistics Management

- Analysis of LAV Depot Maintenance
- Army LOG MOD
- ASDS Product Support Analysis
- Cold-chain Logistics
- Contractors Supporting Military Operations
- Diffusion/Variability on Vendor Performance Evaluation
- Evolutionary Acquisition
- Lean Six Sigma to Reduce Costs and Improve Readiness
- Naval Aviation Maintenance and Process Improvement (2)
- Optimizing CIWS Lifecycle Support (LCS)
- Outsourcing the Pearl Harbor MK-48 Intermediate Maintenance Activity
- Pallet Management System
- PBL (4)
- Privatization-NOSL/NAWCI
- RFID (6)

- Risk Analysis for Performance-based Logistics
- R-TOC AEGIS Microwave Power Tubes
- Sense-and-Respond Logistics Network
- Strategic Sourcing

Program Management

- Building Collaborative Capacity
- Business Process Reengineering (BPR) for LCS Mission Module Acquisition
- Collaborative IT Tools Leveraging Competence
- Contractor vs. Organic Support
- Knowledge, Responsibilities and Decision Rights in MDAPs
- KVA Applied to AEGIS and SSDS
- Managing the Service Supply Chain
- Measuring Uncertainty in Earned Value
- Organizational Modeling and Simulation
- Public-Private Partnership
- Terminating Your Own Program
- Utilizing Collaborative and Three-dimensional Imaging Technology

A complete listing and electronic copies of published research are available on our website: www.acquisitionresearch.org

THIS PAGE INTENTIONALLY LEFT BLANK

ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL
555 DYER ROAD, INGERSOLL HALL
MONTEREY, CALIFORNIA 93943

www.acquisitionresearch.org

Program-Awareness via Lexical Link Analysis (LLA)

Dr. Ying Zhao, Dr. Shelley P. Gallup, Dr. Douglas J. MacKinnon

Research Associate Professors, Distributed Information Systems Experimentation, Naval Postgraduate School

Research Questions

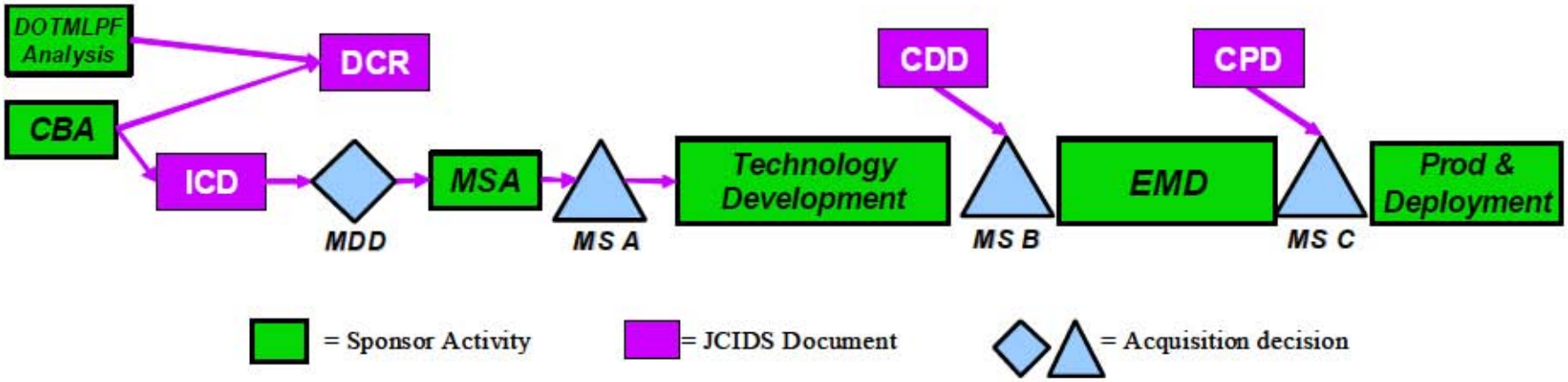
Conceptual: 1) Can the information that emerges from the acquisition process be used to produce overall *awareness* of the *fit* between programs/projects/systems and *needs* for which they were intended?. 2) If a higher level of *awareness* is possible, will that enable system level regulation of programs/projects/systems for improvement of the acquisition system?

Focused: 1) Based on the normal evolution of documentation and current data-based program information, can requirements (needs) be connected to system capabilities? 2) Can requirements gaps be revealed?

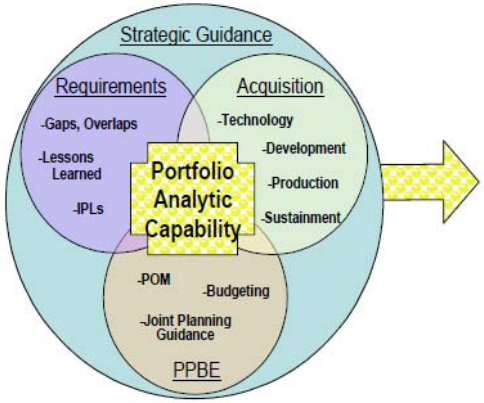
Theory development: Is there a correlation between system interdependency (links/relationships) and development costs?

Methodology: Is it possible to use natural language and other documentation (roughly, unformatted data) to produce visualization of the internal constructs useful for management, through lexical link analysis (LLA)?

Critical Need: Automation



JCIDS Process and Acquisition Decisions
(From J-8 CJCSI 3170.01G)(JCIDS, 2009)



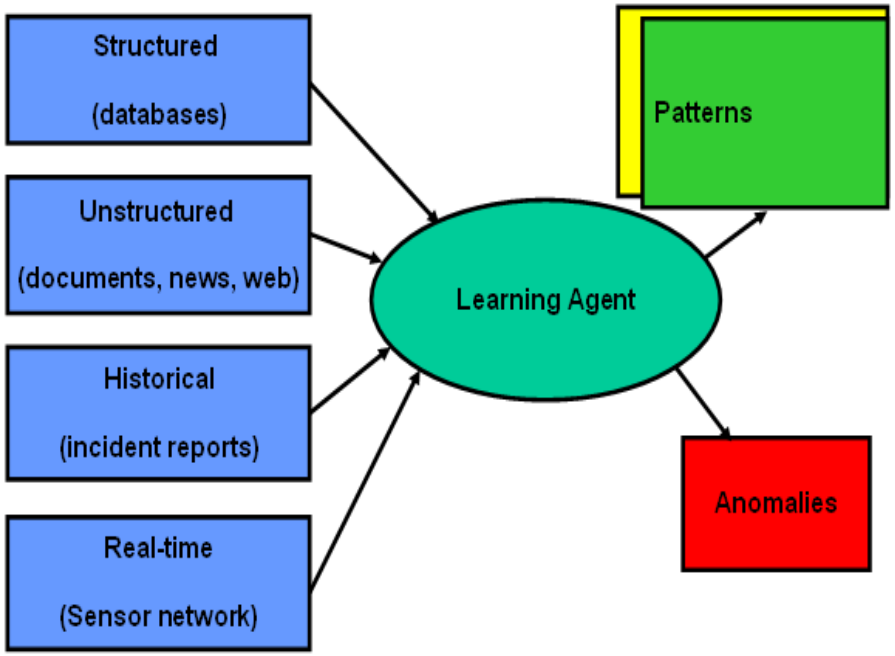
Multiple Portfolio Views:
- Systems vs. Capabilities
- Investment vs. Capabilities
- System Context
- Highly dependent programs (Joint Enablers)
- Procurement Optimization
- S&T vs. future needs
- Sustainment Efficiency
- Market Value

- Data is too voluminous, unformatted and unstructured!
- Need automation
  - Extract relations among PE, MDAP and ACATII
  - Extract costs

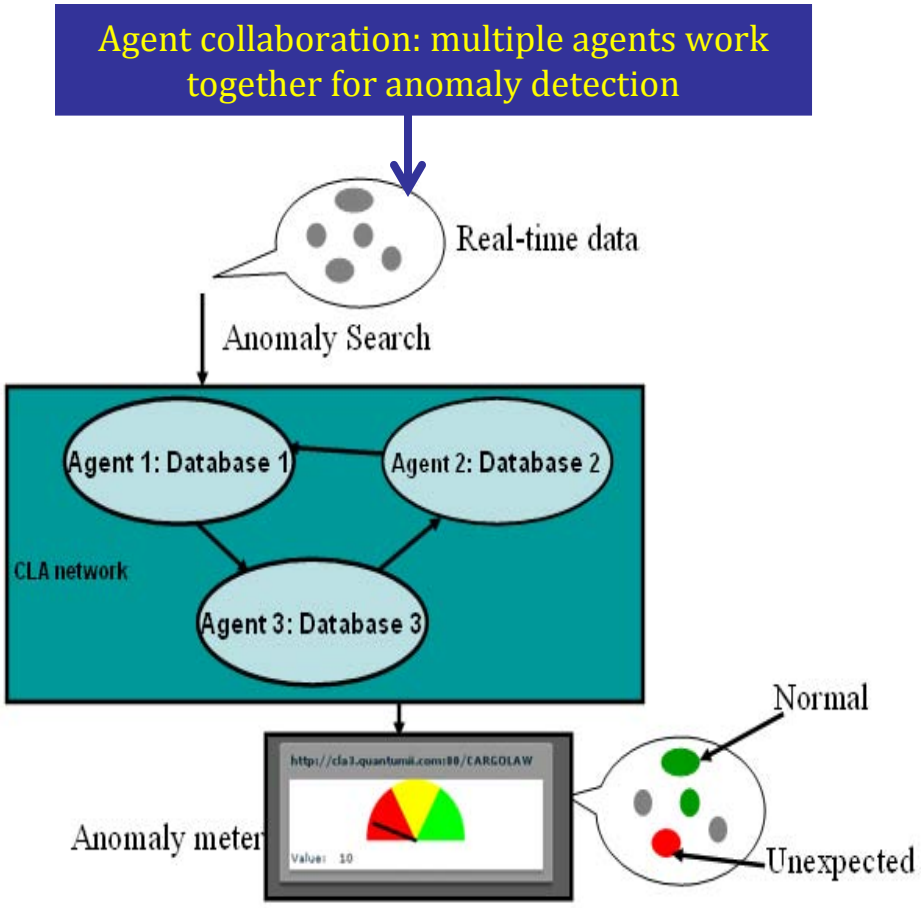
LLA for Analysis of Unstructured Data

- Apply Collaborative Learning Agents
  - Separate patterns and anomalies
  - Parallel computing using NPS High Performance Center (HPC)
- Develop Visualization
  - AutoMap
  - Radar
  - Matrix
- Conduct Pre-processing Steps
  - Named Entity Extraction
    - Leave out people, places and organizations
  - Parts of Speech Tagging
    - Separate nouns, verbs, adjectives, adverbs

Apply Learning Agents to Perform LLA



A learning agent ingests structured, unstructured, historical or real-time data and separate patterns and anomalies.



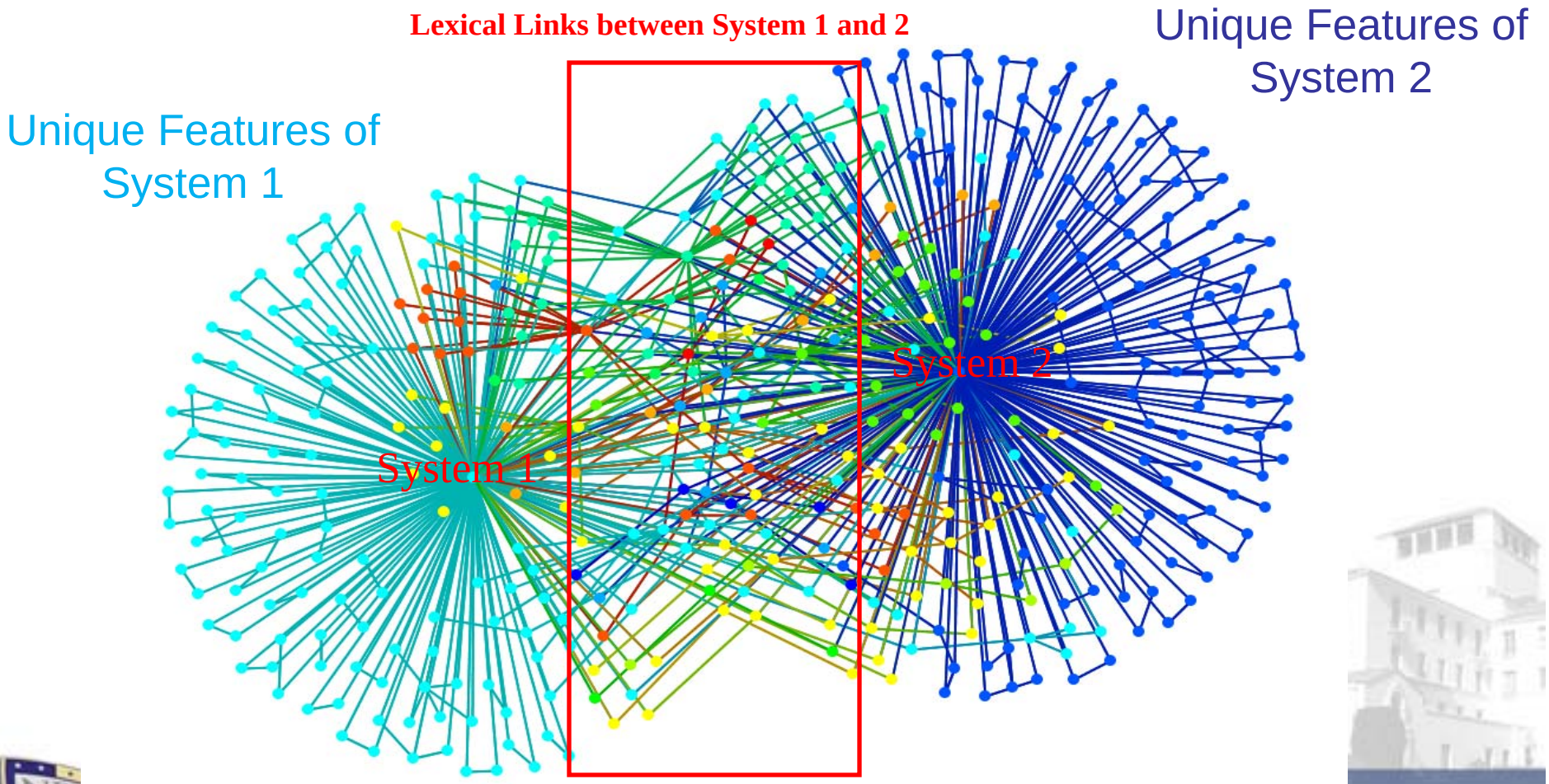
What is a learning agent?

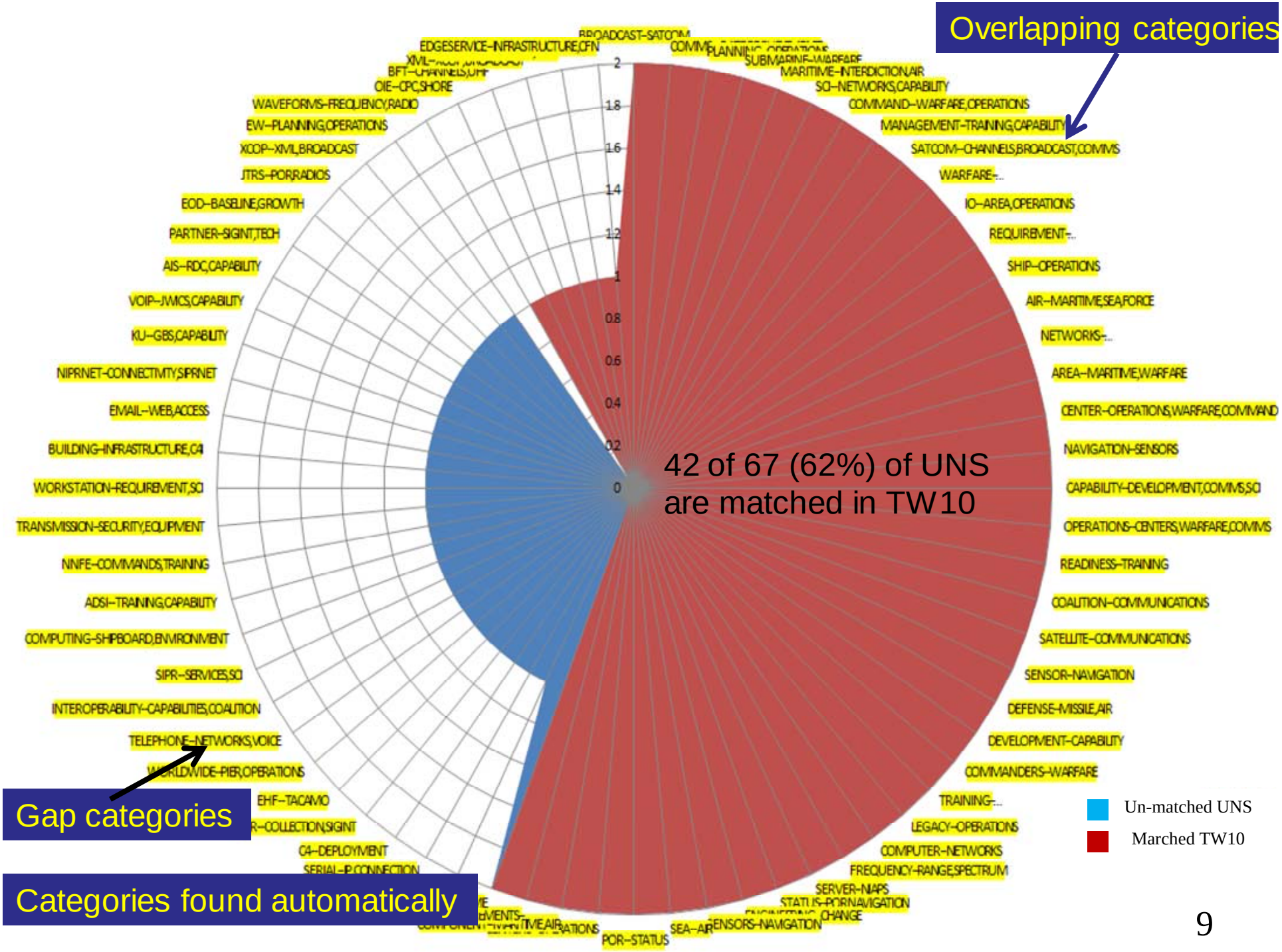
- A computer program or software
  - Installed in a computer with permission
  - Perform automatic tasks
- Multi-agent, distributed networks are capable of
  - Self-managing (Hinchey et al, 2006)
  - Self-healing (Dashofy et al, 2002)
  - Self-optimizing, self-configuring, self-adapting…
- Our learning agent
  - Related to
    - Reinforcement learning (Sutton 1998)
    - Bayesian belief networks (Pearl, 1986; Ben-Gal, 2007)
    - Hidden Markov Models (Huang 1990)
  - Learning patterns and anomalies

Example
Analysis of Urgent Need Statements (UNS)

- Analyzed three lists of classified needs statements and links to Trident Warrior 10 technology capabilities
  - Navy classified UNS
  - C5F (5th Fleet)
  - Integrated Priority List (IPL)
    - CENTCOM and NAVCENT
- Validity checked by Subject Matter Experts

Visualization



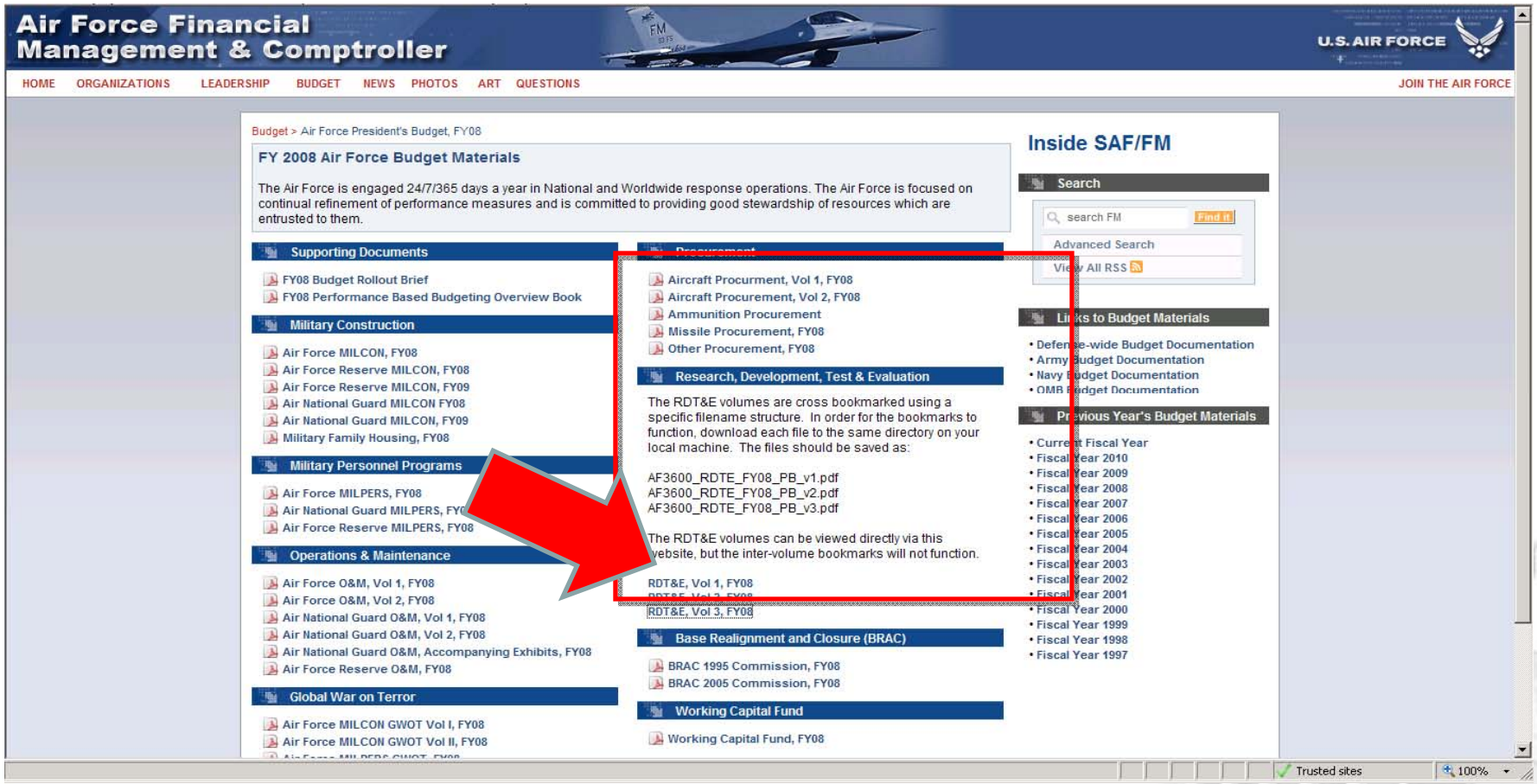

Overlapping categories
42 of 67 (62%) of UNS are matched in TW10
Gap categories
Categories found automatically
Un-matched UNS
Marched TW10
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
EDGESERVICE-INFRASTRUCTURE,CFN
BFT--CHANNELS,UHF
OIE--CPC,SHORE
WAVEFORMS-FREQUENCY,RADIO
EW--PLANNING,OPERATIONS
XCOP--XIVIL,BROADCAST
JTRS--POR,RADIOS
EOD--BASELINE,GROWTH
PARTNER-SIGINT,TECH
AIS--RDC,CAPABILITY
VOIP--JWICS,CAPABILITY
KU--GBS,CAPABILITY
NIPRNET-CONNECTIVITY,SIPRNET
EMAIL--WEB,ACCESS
BUILDING-INFRASTRUCTURE,C4
WORKSTATION-REQUIREMENT,SCI
TRANSMISSION-SECURITY,EQUIPMENT
NNFE--COMMANDS,TRAINING
ADSI--TRAINING,CAPABILITY
COMPUTING-SHIPBOARD,ENVIRONMENT
SIPR--SERVICES,SCI
INTEROPERABILITY-CAPABILITIES,COALITION
TELEPHONE-NETWORKS,VOICE
WORLDWIDE-PIER,OPERATIONS
EHF--TACAMO
C4--DEPLOYMENT
SERIAL-IP,CONNECTION
POR--STATUS
SEA--AIR
SENSORS-NAVIGATION
SERVER-NIAPS
COMPUTER-NETWORKS
FREQUENCY-RANGE,SPECTRUM
LEGACY--OPERATIONS
TRAINING--...
COMMANDERS-WARFARE
DEVELOPMENT-CAPABILITY
DEFENSE-MISSILE,AIR
SENSOR-NAVIGATION
SATELLITE--COMMUNICATIONS
COALITION--COMMUNICATIONS
READINESS--TRAINING
OPERATIONS-CENTERS,WARFARE,COMMS
CAPABILITY-DEVELOPMENT,COMMS,SCI
NAVIGATION-SENSORS
CENTER--OPERATIONS,WARFARE,COMMAND
AREA--MARITIME,WARFARE
NETWORKS-...
AIR--MARITIME,SEA,FORCE
SHIP--OPERATIONS
REQUIREMENT-...
IO--AREA,OPERATIONS
WARFARE-...
SATCOM--CHANNELS,BROADCAST,COMMS
MANAGEMENT-TRAINING,CAPABILITY
COMMAND--WARFARE,OPERATIONS
SCI--NETWORKS,CAPABILITY
MARITIME--INTERDICTION,AIR
SUBMARINE-WARFARE
PLANNING,OPERATIONS
BROADCAST-SATCOM

Large Scale Data

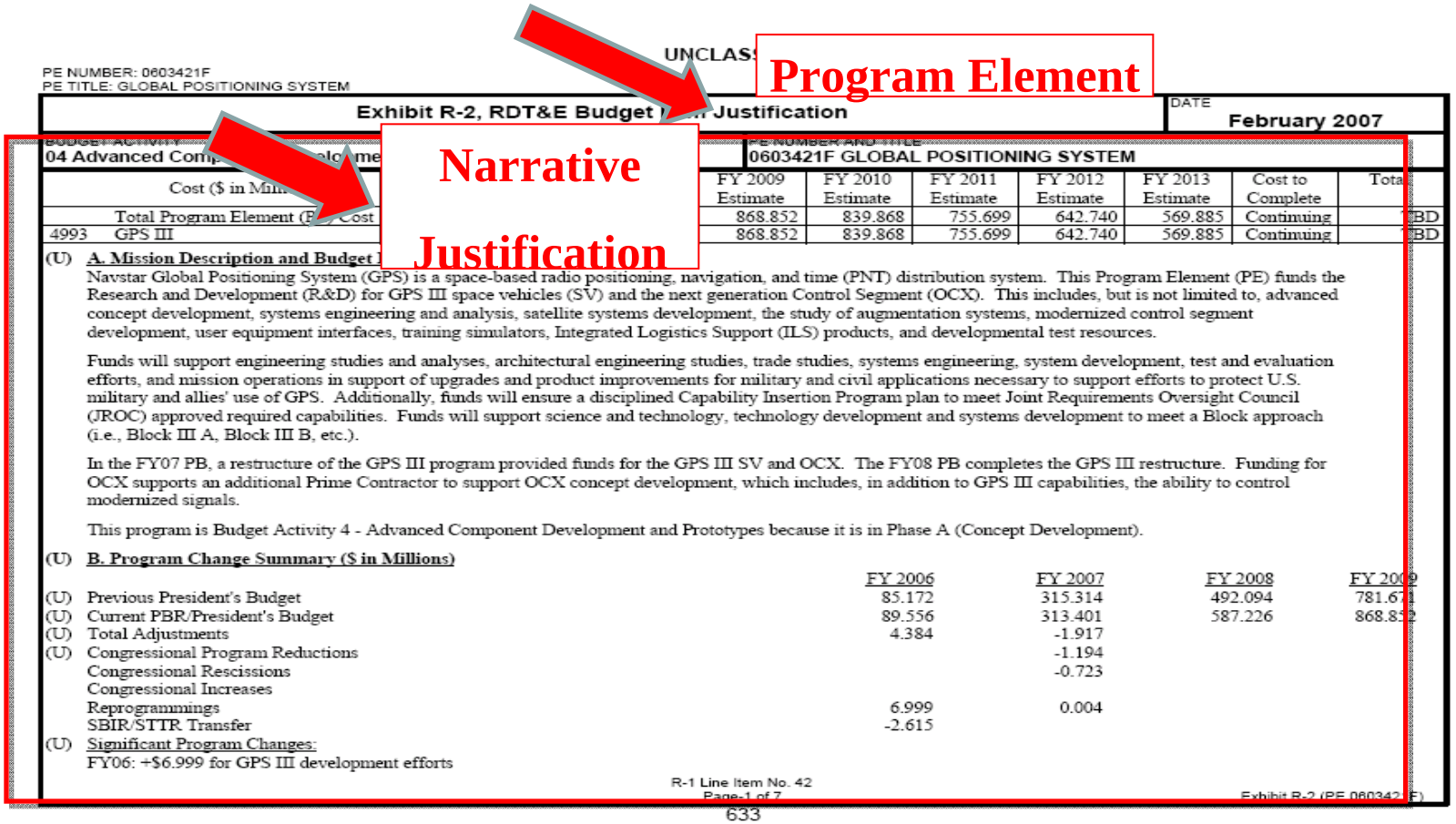
- Program Elements
- Programs
  - Major DOD Acquisition Programs (MDAP)
  - Acquisition Category II (ACAT II)
- UJTLS
- Source
  - Rob Flowe OUSD(AT&L)/ARA/EI

DOD Program Elements

http://comptroller.defense.gov/defbudget/fy2009/index.html



PE Narrative Justification



UNCLAS

PE NUMBER: 0603421F
PE TITLE: GLOBAL POSITIONING SYSTEM

Program Element

Exhibit R-2, RDT&E Budget [illegible] Justification

DATE
February 2007

Narrative Justification

04 Advanced Comp[illegible]

PE NUMBER AND TITLE
0603421F GLOBAL POSITIONING SYSTEM

Cost ($ in Mil[illegible]		FY 2009 Estimate	FY 2010 Estimate	FY 2011 Estimate	FY 2012 Estimate	FY 2013 Estimate	Cost to Complete	Tota
Total Program Element (P[illegible] Cost		868.852	839.868	755.699	642.740	569.885	Continuing	TBD
4993	GPS III	868.852	839.868	755.699	642.740	569.885	Continuing	TBD

(U) **A. Mission Description and Budget [illegible]**

Navstar Global Positioning System (GPS) is a space-based radio positioning, navigation, and time (PNT) distribution system. This Program Element (PE) funds the Research and Development (R&D) for GPS III space vehicles (SV) and the next generation Control Segment (OCX). This includes, but is not limited to, advanced concept development, systems engineering and analysis, satellite systems development, the study of augmentation systems, modernized control segment development, user equipment interfaces, training simulators, Integrated Logistics Support (ILS) products, and developmental test resources.

Funds will support engineering studies and analyses, architectural engineering studies, trade studies, systems engineering, system development, test and evaluation efforts, and mission operations in support of upgrades and product improvements for military and civil applications necessary to support efforts to protect U.S. military and allies' use of GPS. Additionally, funds will ensure a disciplined Capability Insertion Program plan to meet Joint Requirements Oversight Council (JROC) approved required capabilities. Funds will support science and technology, technology development and systems development to meet a Block approach (i.e., Block III A, Block III B, etc.).

In the FY07 PB, a restructure of the GPS III program provided funds for the GPS III SV and OCX. The FY08 PB completes the GPS III restructure. Funding for OCX supports an additional Prime Contractor to support OCX concept development, which includes, in addition to GPS III capabilities, the ability to control modernized signals.

This program is Budget Activity 4 - Advanced Component Development and Prototypes because it is in Phase A (Concept Development).

(U) **B. Program Change Summary ($ in Millions)**

		FY 2006	FY 2007	FY 2008	FY 2009
(U)	Previous President's Budget	85.172	315.314	492.094	781.67[illegible]
(U)	Current PBR/President's Budget	89.556	313.401	587.226	868.85[illegible]
(U)	Total Adjustments	4.384	-1.917		
(U)	Congressional Program Reductions		-1.194		
	Congressional Rescissions		-0.723		
	Congressional Increases				
	Reprogrammings	6.999	0.004		
	SBIR/STTR Transfer	-2.615			
(U)	Significant Program Changes:				

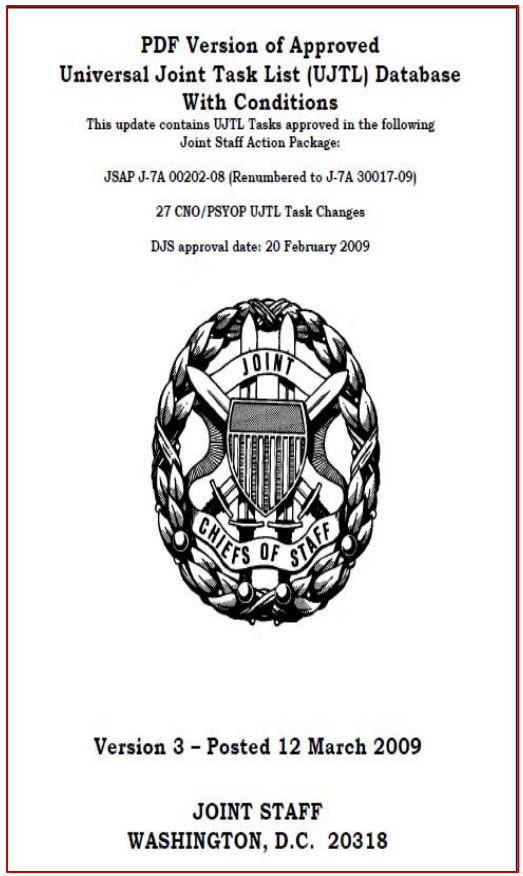
FY06: +$6.999 for GPS III development efforts

R-1 Line Item No. 42
Page-1 of 7
Exhibit R-2 (PE 060342[illegible]F)

633

UNCLASSIFIED

Acquisition Documents


PDF Version of Approved
Universal Joint Task List (UJTL) Database
With Conditions
This update contains UJTL Tasks approved in the following
Joint Staff Action Package:

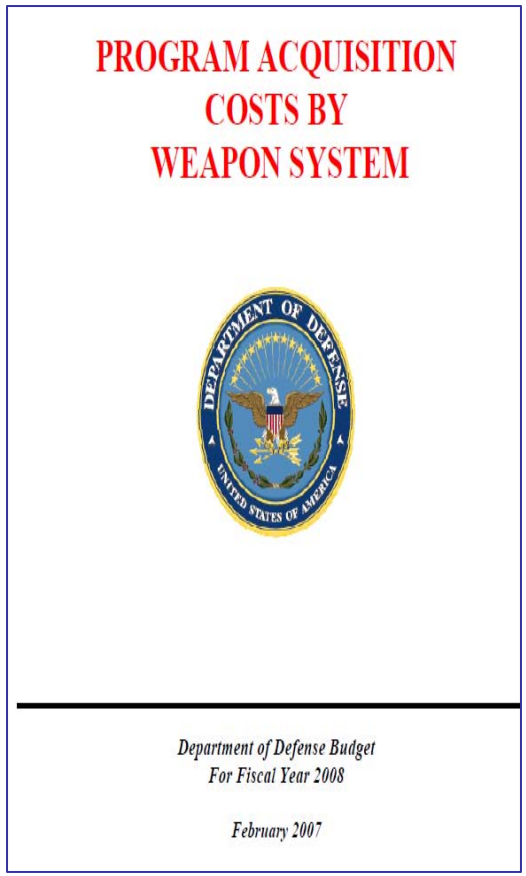
JSAP J-7A 00202-08 (Renumbered to J-7A 30017-09)

27 CNO/PSYOP UJTL Task Changes

DJS approval date: 20 February 2009

Version 3 – Posted 12 March 2009

JOINT STAFF
WASHINGTON, D.C. 20318

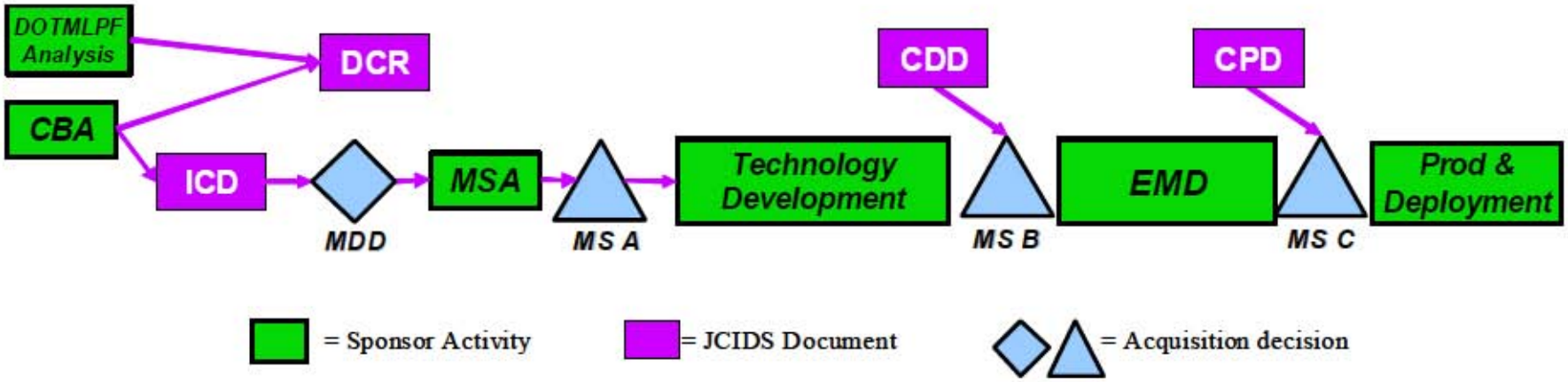

PROGRAM ACQUISITION
COSTS BY
WEAPON SYSTEM

Department of Defense Budget
For Fiscal Year 2008

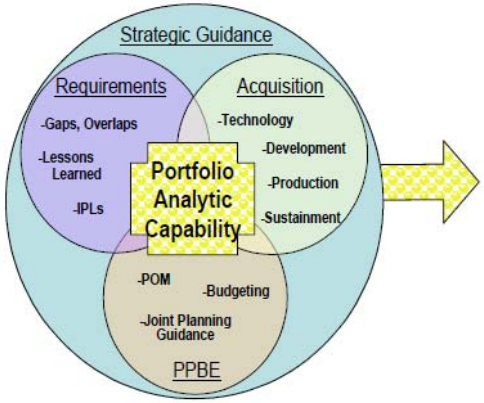
February 2007

- Program Elements: RDTE books
  - Air Force
  - Army
    - http://asafm.army.mil/Document.aspx?OfficeCode=1200
  - Navy
    - http://www.finance.hq.navy.mil/fmb/11pres/BOOKS.htm
- Universal Joint Task List
- Weapon Books

Critical Need: Automation



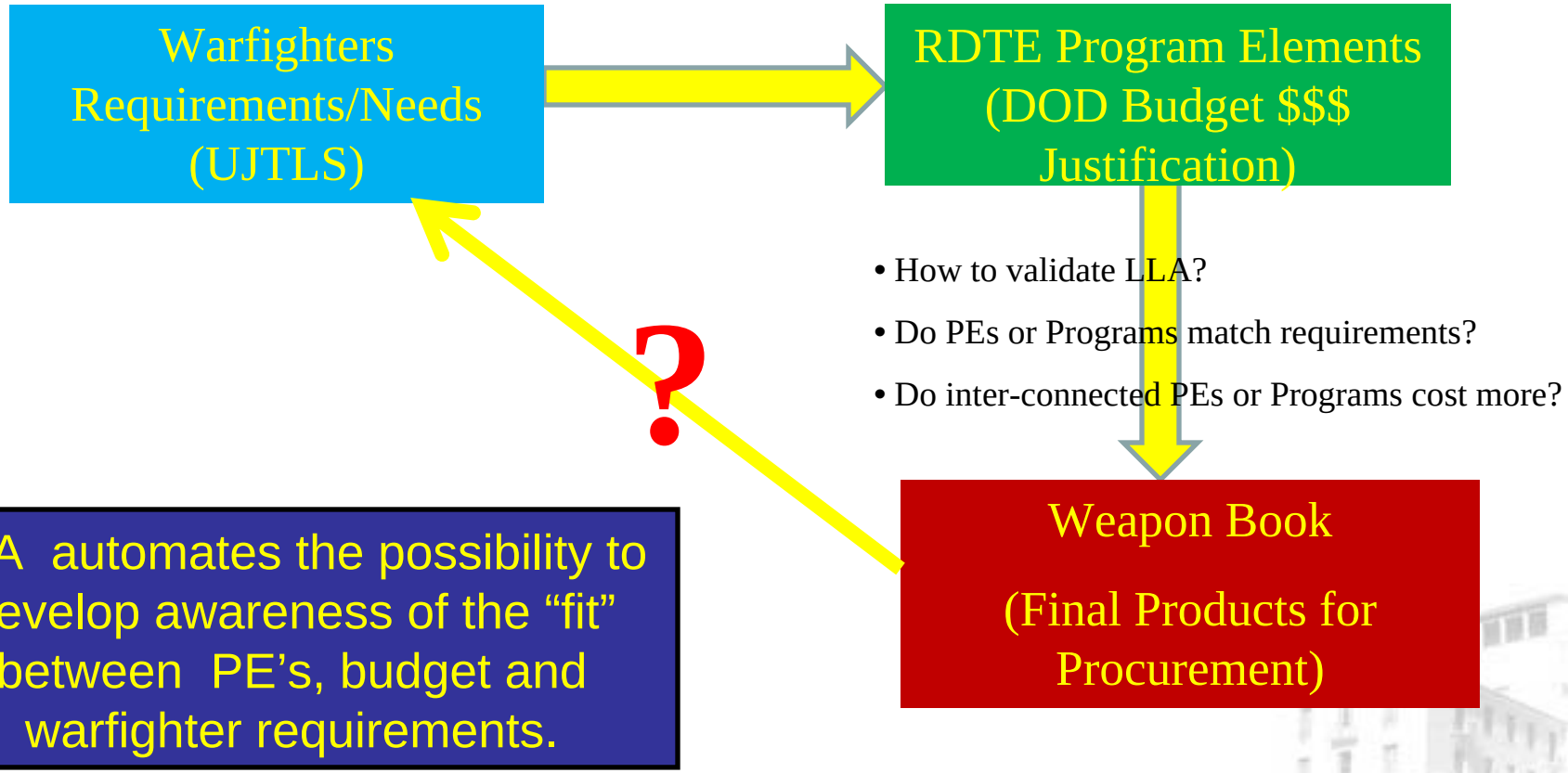
JCIDS Process and Acquisition Decisions
(From J-8 CJCSI 3170.01G)(JCIDS, 2009)



Multiple Portfolio Views:
- Systems vs. Capabilities
- Investment vs. Capabilities
- System Context
- Highly dependent programs (Joint Enablers)
- Procurement Optimization
- S&T vs. future needs
- Sustainment Efficiency
- Market Value

- Data is too voluminous, unformatted and unstructured!
- Need automation
  - Extract relations among PE, MDAP and ACATII
  - Extract costs

LLA Methodology Can Help!



LLA automates the possibility to develop awareness of the "fit" between PE's, budget and warfighter requirements.

PE Links Identified by Human Analysts (Used for LLA Validation)

UNCLASSIFI

Exhibit R-2a, RDT&E Project Justifica

DATE: May 2009

BUDGET ACTIVITY: 05 System Development and Demonstration (SDD)

PE N R AND TITLE: 0604602F Armament/Ordnance Development

PROJECT NUMBER AND TITLE: 5361 Stores-Aircraft Interface

Cost ($ in Millions)	FY 2008 Actual	FY 2009 Estimate	FY 2010 Estimate	FY Esti	FY 2012 Estimate	FY 2013 Estimate	FY 2014 Estimate	FY 2015 Estimate	Cost to Complete	Total
5361 Stores-Aircraft Interface	0.000	0.000	6.685		0.000	0.000	0.000	0.000	Continuing	TBD
Quantity of RDT&E Articles	0	0	0	0	0	0	0	0		

In FY 2010, Project 5361, Stores-Aircraft Interface (new), efforts were transferred from PE 0605011F, RDT&E for Aging Aircraft, Project 654685, Universal Armament Interface (UAI), in order to properly fund the maturing technology.

(U) **A. Mission Description and Budget Item Justification**

Universal Armament Interface (UAI) is an Air Force initiative to develop, enhance, and implement standardized interfaces in aircraft, weapons and mission planning to support integration of weapons independent of aircraft Operation Flight Program (OFP) cycles. UAI is currently being implemented on the F-15E and F-16 Block 40/50 aircraft, Small Diameter Bomb (SDB) I and II, Joint Direct Attack Munition (JDAM), Joint Air-to-Surface Stand-off Missile (JASSM) and Precision Guided Munitions Planning Software (PGMPS). Additional aircraft and weapons have program plans to implement UAI. The UAI program office is responsible for development and enhancement of the standard, provision of certification tools (test assets) and implementation support to aircraft and weapons.

The UAI efforts were transferred (1) to ensure continued funding for UAI through the FYDP (PE 0605011F will be zeroed out in FY 2010 due to higher Air Force priorities), and (2) to properly fund the maturing technology. The new project number is established to provide greater visibility into UAI's budget. Funding UAI via the Arm/Ord PE will ensure that platform and weapon program offices have the support required to implement and update UAI.

This program is in Budget Activity 5 - System Development and Demonstration (SDD) because it supports armament integration, an SDD-type activity.

(U) B. Accomplishments/Planned Program ($ in Millions)	FY 2008	FY 2009	FY 2010
(U) ICD Dev/Updates			5.702
(U) UAI Common Component			0.786
(U) Certification Tool			0.197
(U) Total Cost	0.000	0.000	6.685

This is not a new start; these efforts were performed under PE 0605011F, RDT&E for Aging Aircraft, in FY 2008 and FY 2009.

0604602F references 0605011F Forward Link

0605011F referenced by 0604602F Backward Link

(U) **C. Other Program Funding Summary ($ in Millions)**

FY 2008 Actual	FY 2009 Estimate	FY 2010 Estimate	FY 2011 Estimate	FY 2012 Estimate	FY 2013 Estimate	FY 2014 Estimate	FY 2015 Estimate	Cost to Complete	Total Cost

(U) N/A

(U) **D. Acquisition Strategy**

In December 2004, under the authority of a class Justification and Approval (J&A), the UAI program office awarded individual Cost Plus Fixed Fee (CPFF) contracts to Boeing, Lockheed-Martin, Northrop-Grumman and Raytheon. These four vendors are the Original Equipment Manufacturers (OEMs) for approximately 90% of the Department of Defense' platforms and weapons. Each OEM is responsible for a different piece of the total UAI requirement based on its platform or weapon expertise.

Project 5361 | R-1 Line Item No. 77 Page-9 of 13 | Exhibit R-2a (PE 0604602F)

469

UNCLASSIFIED

Validate LLA and Discover Statistically Significant Correlation

PEs

	0101113F	0101122F	0101221N	0101226N	0101313F
0204413N	36,PROFILE,CONTROLLER,ARTICLES,TACTICAL,FUNCTIONAL,TRANSFERS,DIGITAL	D,BLANK,PERFORMED,INTENTIONALLY,ELECTRICAL,ARTICLES,FUNCTIONAL,ACCO	D,BLANK,SUPT,NAVY,TRANSITION,INTENTIONALLY,DIGITAL,CONTRACTS,ARTICLES	S,INITIATIVES,METRICS,SUBTOTAL,TRANSFERS,TOTALS,PRIOR,READINESS,CATEG	TIONALLY,CONTRVES,TACTICAL
0204571N	NERSHIP,DEVICES,OBSOLESCENCE,NAVIGATION,SUPPORTED,DAMAGE,IDENTIFIE	VIGATION,IDENTIFIED,PROFILE,ALTERNATIVES,UTILIZING,STRIKE,MODIFICATIONS	LEET,CYCLE,DEVICES,OBSOLESCENCE,NAVIGATION,IDENTIFIED,COMMENCE,SUPP	E,UNCLASSIFIED,LOOP,COUNTERMEASURE,FACILITY,ENGINEERING,MILESTONE,RE	TION,DEFINED,LE,EVALUATIONS,
0204574N	TION,REPORTING,INTEGRATED,MANUFACTURING,ENHANCEMENTS,DEVELOPS,AR	ANK,INTENTIONALLY,ARTICLES,INVENTORY,PRIOR	EMERGENT,CERTIFICATION,UNCLASSIFIED,TESTED,INTEGRATED,SUBSURFACE,BLA	ELOPS,ARTICLES,METRICS,SUBTOTAL,TRANSFERS,DEC,TOTALS,PRIOR,CATEGORY,Q	PLOYMENT,INTEGENTATION,STRAT
LLA:# of Matched Word Hubs	261	88	413	54	
LLA: Overall Match Score	156125	63013	326240	32278	
LLA:# of Unique Word Hubs					
PE Forward Links	1				
PE Backward Links	1		1		
PE Links(Forward+Backward)	2	0	1	0	
2009 Cost	38651	936	84120	7364	
	0.396594525				

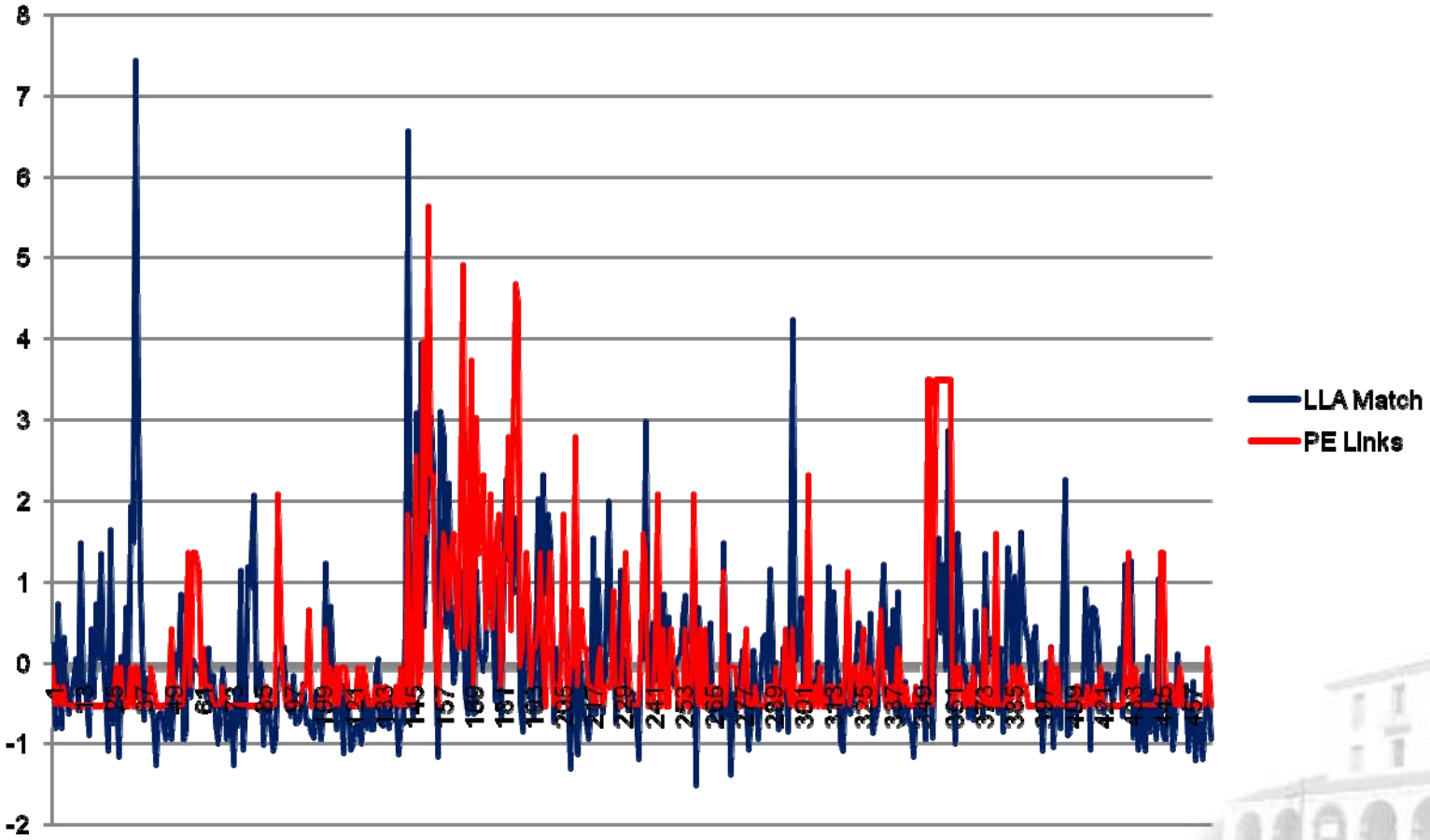
PEs

From LLA using the narrative descriptions of each PE

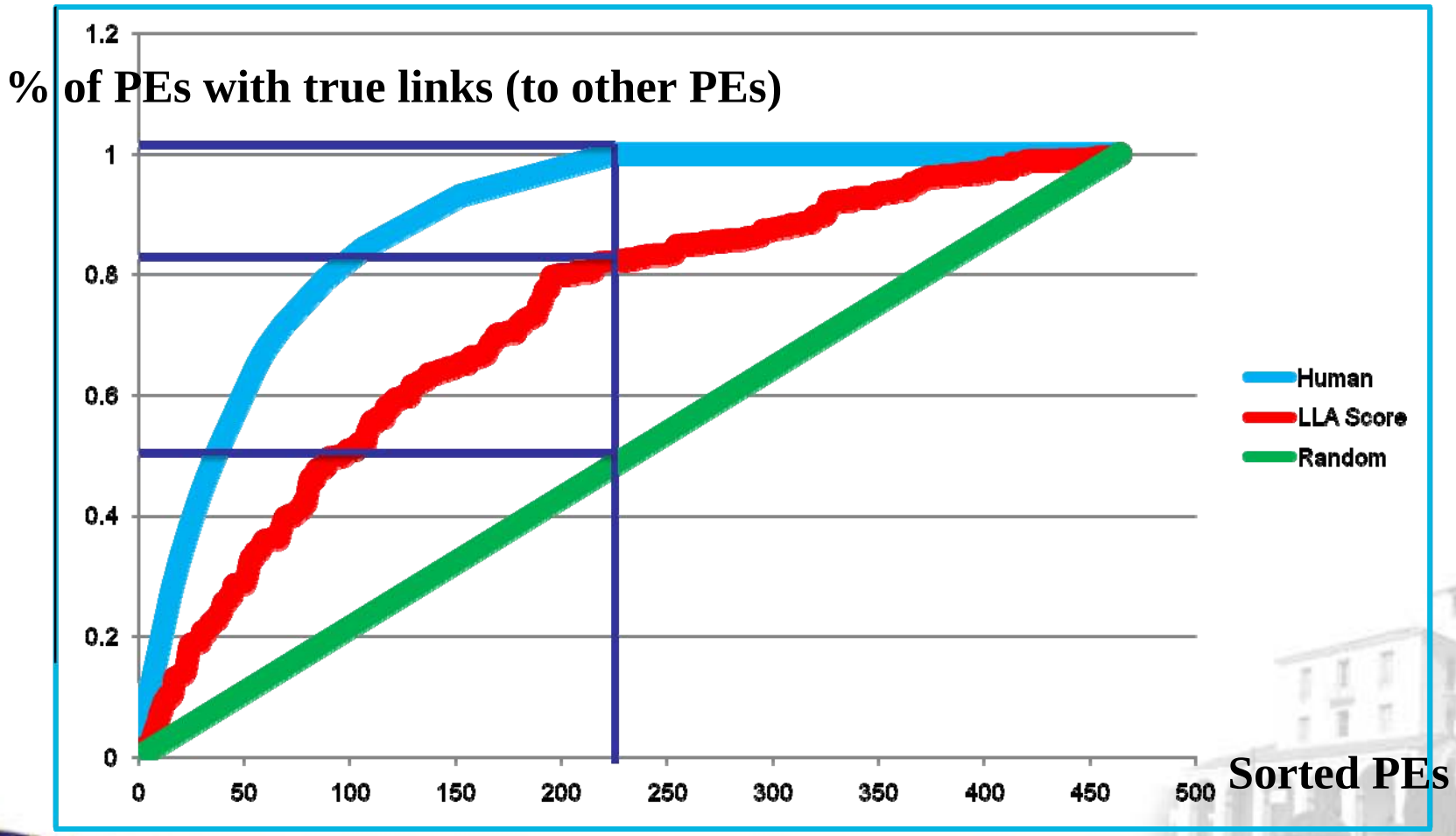
Pearson correlation between the two is 0.39 (p-value=0.0000001)

From human analysts

Visualize the Correlation



Use LLA Scores to Predict PE Links: Gains Chart



pe_pe.matrix.html - Microsoft Excel

	A	B	C
1		0101113F.txt	0101122F.txt
2	0604226F.txt	0027656.27;STERLING--VA;OWNERSHIP--COST,COSTS;BP16--INITIAL,PE	5.33
3	0101126F.txt	0019881.86;OWNERSHIP--COST,COSTS;BP16--INITIAL,PE	8.1
4	0207581F.txt	0018671.22;BP16--INITIAL,PE	5.72
5	0603235N.txt	0018667.64;SOURCED--DATA,SOFTWARE	3.16
6	0302015F.txt	0017172.55;OGDEN--AIR,AFB;REPLACES--CURRENT;DEPENDENT--SURVEILLANCE	6.87
7	0207136F.txt	0013337.67;AFMSS--UPGRADES,SS	6.79
8	0207417F.txt	0007315.54;RNP--GLOBAL,SURVEILLANCE;GWOT--FUNDING	6.69
9	0207249F.txt	0006227.37;ATP--EFFORT,REQUIREMENTS	7.9
10	0401119F.txt	0006133.00;OWNERSHIP--COST,COSTS;WARTIME--CAPABILITY,MISSIONS	7.14
11	0207590F.txt	0004917.94;LITENING--INTEGRATION,TARGETING	7.96
12	0204229N.txt	0004916.45;WARTIME--CAPABILITY,MISSIONS	4.91
13	0303601F.txt	0004548.71;FAB--INCREMENT[illegible]REMELY--FREQUENCY	7.79
14	0602271N.txt	0004227.58;EXTREMELY--FREQ[illegible]	3.22
15	0604503N.txt	0004227.12;EXTREMELY--FREQUENC	3.79
16	0401219F.txt	0003843.68;REPLACE	
17	0303109N.txt	0003807.47;EXTREME	
18	0901212F.txt	0003596.23;NORMAL	
19	0605709A.txt	0003592.63;NORMAL	
20	0602236N.txt	0002746.67;EXTREME	
21	0205633N.txt	0002698.57;OWNERS	
22	0604567N.txt	0002697.52;OWNERS	
23	0603635M.txt	0002697.32;OWNERS	
24	0204163N.txt	0002596.36;EXTREME	
25	0605976F.txt	0002396.59;GWOT--	
26	0605805A.txt	0002288.60;OWNERS	
27	0305114F.txt	0002081.56;DEPENDI	
28	0604633A.txt	0002079.14;DEPENDI	
29	0401218F.txt	0001767.75;REPLACE	
30	0101127F.txt	0001538.71;EXTREME	
31	0603430F.txt	0001537.24;EXTREME	
32	0604240F.txt	0001536.97;EXTREME	
33	0303131F.txt	0001536.38;EXTREME	
34	0603432F.txt	0001536.18;EXTREME	
35	0603854F.txt	0001535.96;EXTREME	
36	0602235N.txt	0001535.65;EXTREME	
37	0605712F.txt	0001534.91;EXTREME	

Links discovered by LLA

Links noted by analysts

UNCLASSIFIED

Exhibit R-2, RDT&E Budget Item Justification: PB 2011 Air Force	**DATE:** February 2010
APPROPRIATION/BUDGET ACTIVITY 3600: *Research, Development, Test & Evaluation, Air Force* BA 7: *Operational Systems Development*	**R-1 ITEM NOMENCLATURE** PE 0101113F: *B-52 SQUADRONS*

The B-52 Extremely High Frequency (EHF) will integrate and install the B-52 fleet with assured and survivable two-way EHF SATCOM link for Emergency Action Messages (EAMs) and report-backs to meet Joint Chiefs of Staff (JCS) nuclear protected Information Exchange Requirements (IER). The B-52 EHF will integrate the Family of Advanced Beyond-Line-of-Sight (BLOS) Terminal (FAB-T) Increment 1 system developed and procured by Space and Missile Command (SMC) through PE 0303601F. The FAB-T system consists of the Operator Interface Group, Modem Processor Group, and Antenna Group. The B-52 EHF will integrate the following capability into the CONECT baseline B-52 architecture: a high data rate BLOS communication link supporting IP-based Global Information Grid (GIG) interoperability. The two Multi-function Color Displays (MFCDs) and the additional J-Series Messages that were to be integrated into CONECT with the B-52 EHF have been moved to the Strategic Radar Replacement and CONECT programs, respectfully. In addition, the automated reporting of aircraft fuel level status off-board the jet capability will also be moved from the B-52 EHF. Disposition of this capability is pending an approved acquisition strategy. The B-52 EHF program is planned to be accomplished in three increments. Increment 1 is the up front program planning and risk reduction trade studies on items like radome mounting, environmental cooling system (ECS) capabilities, antenna boresighting, etc. Increment 2 will integrate, and install the FAB-T equipment for strategic connectivity, as well as implement trade study solutions. In addition, the ECS will need to be upgraded or replaced. The ECS modification requirements will allow enough margin to accommodate near-term, future roadmap efforts. Finally, Increment 3 will provide GIG and net ready capability as well as full integration with other B-52 systems.

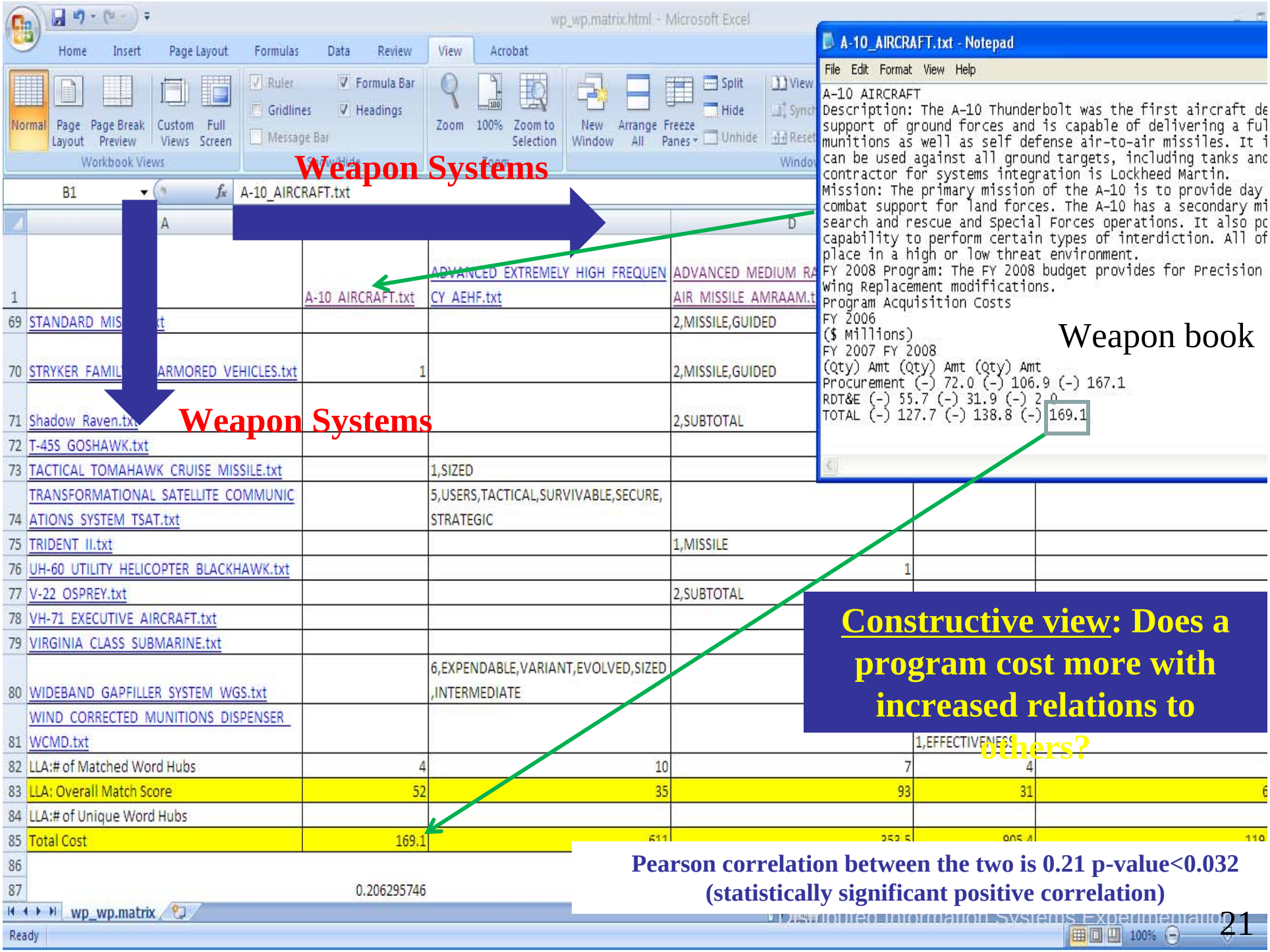
Trainers and upgrades for CONECT & EHF

In order to maintain currency with the latest aircraft configuration, the CONECT and EHF programs will update existing trainers or use computer-based training to add CONECT and EHF functionality to meet user-training requirements and establish a system integration laboratory (SIL) for updates of the Weapon System Trainers (WST).

Advanced Targeting Pod Functionality

The B-52 Modernization program fully integrates Advanced Targeting Pods (ATP) by linking pod control, display and target geo-location with the B-52 Offensive Avionics System (OAS). The B-52 ATP effort continues the ATP (Sniper or LITENING) integration effort that began in FY07 with GWOT funding. The ATP effort develops aircraft software updates to add and incorporate advanced pod functionality into the B-52. In addition, this effort upgrades the software functions of the new Alternate Mission Equipment (AME) (Multi Function Display and the Integrated Hand Controller), developed and procured under the B-52 Advanced Weapons Integration (AWI) modification, and enables the B-52 to utilize a LITENING or Sniper pod. This effort provides hardware and software upgrades to the existing aircrew/ maintenance trainers and the SIL.

Weapons Improvements


wp_wp.matrix.html - Microsoft Excel
Weapon Systems
Weapon Systems
A-10_AIRCRAFT.txt - Notepad
File Edit Format View Help
A-10 AIRCRAFT
Description: The A-10 Thunderbolt was the first aircraft de
support of ground forces and is capable of delivering a ful
munitions as well as self defense air-to-air missiles. It i
can be used against all ground targets, including tanks anc
contractor for systems integration is Lockheed Martin.
Mission: The primary mission of the A-10 is to provide day
combat support for land forces. The A-10 has a secondary mi
search and rescue and Special Forces operations. It also pc
capability to perform certain types of interdiction. All of
place in a high or low threat environment.
FY 2008 Program: The FY 2008 budget provides for Precision
wing Replacement modifications.
Program Acquisition Costs
FY 2006
($ Millions)
FY 2007 FY 2008
(Qty) Amt (Qty) Amt (Qty) Amt
Procurement (-) 72.0 (-) 106.9 (-) 167.1
RDT&E (-) 55.7 (-) 31.9 (-) 2.0
TOTAL (-) 127.7 (-) 138.8 (-) 169.1
Weapon book
Constructive view: Does a program cost more with increased relations to others?
Pearson correlation between the two is 0.21 p-value<0.032
(statistically significant positive correlation)
wp_wp.matrix
Ready

There is a statistically insignificant correlation between weapon systems' RDT&E cost and # of lexical links to ACAT II systems

Weapon Systems

ACAT II Sys

	A	B	C	D	E	
1		A-10 AIRCRAFT.txt	EMELY HIGH FREQUEN CY AEHF.txt	ADVANCED MEDIUM RANGE AIR-TO-AIR MISSILE AMRAAM.txt	AH-64 APACHE.txt	AIR I
11	Unit Water Pod Sy (Camel).pdf	0.04	0.11,VEHICLE,TACTICAL	0.11,MEDIUM	0.05,INTEGRATED	
12	Warfighter Inform Network-Tactical (WIN-T).p	0.22,ENVIRONMENT,COMBAT,INTEGRATION,LOCKHEED	0.37,INTEGRATION,SECURE,COMMUNICATIONS,WARFIGHTER,DATA,TACTICAL,LOCKHEED	0.13,ENVIRONMENT,JOINT	0.16,INTEGRATED,SUPPORTS	0.14,E
13	Wire-Guided (TOW) Missiles.pdf	0.20,PRECISION,VEHICLES,RANGE,MISSILES,THREAT	0.16,LAUNCH,PRIME,VEHICLE,ADVANCED,CONTROL,ANTI	0.24,PRIME,GUIDED,MISSILE,AZ,ADVANCED,RANGE,RAYTHEON	0.18,PRIME,MISSILE,AZ,MOUNTED,CONTROL	0.37,L NGE,R
14	XM101 Common Remotely Operated Weapon Station.pdf	0.20,ARMORED,TARGETS,VEHICLES,RANGE	0.03,VEHICLE	0.07,TARGETS,RANGE	0.13,SENSORS	0.07,T
15	XM307.pdf	0.19,ARMORED,TARGETS,COMBAT,INTEGRATION,VEHICLES	0.19,LAUNCH,PRIME,VEHICLE,INTEGRATION,ADVANCED,VARIANT	0.17,TARGETS,PRIME,ADVANCED,RAYTHEON	0.14,PRIME,MOUNTED	0.19,L AYTHE
16	LLA:# of Matched Word Hubs	24	27	19	25	
17	LLA: Overall Match Score	21.76839827	21.52527465	17.7039979	17.48533462	
18	LLA:# of Unique Word Hubs	8				
19	PE Forward Links					
20	PE Backward Links					
21	PE Links(Forward+Backward)	0				
22	RDT&E Cost	2	603.2	41.4	193.7	
23						
24	Pearson correlation	0.181598022				

Pearson correlation between the two is 0.18
p-value<0.055 statistically insignificant

Correlation between Unique # of LLA Word Hubs and Increasing Procurement Cost

Weapon Systems

ACAT II Sys

	A	C	D	E	
1		A-10 AIRCRAFT.txt	ADVANCED EXTREMELY HIGH FREQUENCY AEHF.txt	ADVANCED MEDIUM RANGE AIR-TO-AIR MISSILE AMRAAM.txt	AH-64 APACHE.txt	AIR IN
111	Unit Water Pod S n (Camel).pdf	0.04	0.11,VEHICLE,TACTICAL	0.11,MEDIUM	0.05,INTEGRATED	
112	Warfighter Infor n Network-Tactical (WIN-T)	0.22,ENVIRONMENT,COMBAT,INTEGRATION,LOCKHEED	0.37,INTEGRATION,SECURE,COMMUNICATIONS,WARFIGHTER,DATA,TACTICAL,LOCKHEED	0.13,ENVIRONMENT,JOINT	0.16,INTEGRATED,SUPPORTS	0.14,E
113	Wire-Guided (TOW) Missiles.pdf	0.20,PRECISION,VEHICLES,RANGE,MISSILES,THREAT	0.16,LAUNCH,PRIME,VEHICLE,ADVANCED,CONTROL,ANTI	0.24,PRIME,GUIDED,MISSILE,AZ,ADVANCED,RANGE,RAYTHEON	0.18,PRIME,MISSILE,AZ,MOUNTED,CONTROL	0.37,L NGE,R
114	XM101 Common Remotely Operated Weapon Station.pdf	0.20,ARMORED,TARGETS,VEHICLES,RANGE	0.03,VEHICLE	0.07,TARGETS,RANGE	0.13,SENSORS	0.07,T
115	XM307.pdf	0.19,ARMORED,TARGETS,COMBAT,INTEGRATION,VEHICLES	0.19,LAUNCH,PRIME,VEHICLE,INTEGRATION,ADVANCED,VARIANT	0.17,TARGETS,PRIME,ADVANCED,RAYTHEON	0.14,PRIME,MOUNTED	0.19,L AYTHE
116	LLA:# of Matched Word Hubs	24	27	19	25	
117	LLA: Overall Match Score	21.76839827	21.52527465	17.7039979	17.48533462	
118	LLA:# of Unique Word Hubs	8	14	7	8	
119	PE Forward Links					
120	PE Backward Links					
121	PE Links(Forward+Backward)	0	0	0	0	
122	**Procurement Cost**	**167.1**	**7.8**	**312.1**	**711.7**	

Pearson correlation between the two is 0.34

p-value < 0.001 statistically significant

Results/Conclusions

- Provided an automated tool to surface important aspects among programs
- Proved LLA Validity for automation
  - Adequately models expected human performance but faster
- Demonstrated correlation among relations between programs
  - Cost drivers: Interrelated and Uniqueness
- Discovered statistically significant correlations of Lexical links between MDAP and ACAII, and RDT&E cost

Future Work

- Extract lexical links for applications
  - Continue to explore available acquisition data
    - Extract the cost of MDAP programs (PNO) from the PE documents
    - Compare with SAR (Selected Acquisition Report) as in the MDAP perspective
  - Search for other correlations among other program attributes
  - Identify more dependent variables
    - Diversity metrics
  - Predict program costs using this methodology
    - Cost and cost growth relative to the Milestone B
    - Cascade effect of program costs

Future Work

- Determine methods to leverage the NPS HPC to analyze larger data sets
- Develop improved graphic illustrations of findings
  - 3-D
  - Dynamic
- Provide an automatic LLA service for program self-awareness
  - Enterprise Lexicon Service
  - Meta-Data Registry
- Establish a complex system theory for a cross-domain
  - Law of requisite variety
  - Design Structure Matrix

Program-Awareness via Lexical Link Analysis (LLA)

Dr. Ying Zhao, Dr. Shelley P. Gallup, Dr. Douglas J. MacKinnon

Research Associate Professors, Distributed Information Systems Experimentation, Naval Postgraduate School